

VISUAL LANGUAGE

EDITED BY: Wendy Sandler, Marianne Gullberg and Carol Padden
PUBLISHED IN: Frontiers in Psychology and Frontiers in Communication





frontiers

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714
ISBN 978-2-88963-078-3
DOI 10.3389/978-2-88963-078-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

VISUAL LANGUAGE

Topic Editors:

Wendy Sandler, University of Haifa, Israel

Marianne Gullberg, Lund University, Sweden

Carol Padden, University of California, San Diego, United States



'Visual Language'

image by Shai Davidi, Sign Language Research Lab, University of Haifa.

Traditionally, research on human language has taken speech and written language as the only domains of investigation. However, there is now a wealth of empirical studies documenting visual aspects of language, ranging from rich studies of sign languages, which are self-contained visual language systems, to the field of gesture studies, which examines speech-associated gestures, facial expressions, and other bodily movements related to communicative expressions. But despite this large body of work, sign language and gestures are rarely treated together in theoretical discussions. This volume aims to remedy that by considering both types of visual language jointly in order to transcend (artificial) theoretical divides, and to arrive at a comprehensive account of the human language faculty. This collection seeks to pave the way for an inherently multimodal view of language, in which visible actions of the body play a

crucial role. The 19 papers in this volume address four broad and overlapping topics: (1) the multimodal nature of language; (2) multimodal representation of meaning; (3) multimodal and multichannel prosody; and (4) acquisition and development of visual language in children and adults.

Citation: Sandler, W., Gullberg, M., Padden, C., eds. (2019). Visual language. Lausanne: Frontiers Media. doi: 10.3389/978-2-88963-078-3

Table of Contents

06 Editorial: Visual Language

Wendy Sandler, Marianne Gullberg and Carol Padden

1 THE NATURE OF LANGUAGE AS MULTIMODAL – SPEECH-GESTURE-SIGN

09 Why We Should Study Multimodal Language

Pamela Perniss

14 Language as Description, Indication, and Depiction

Lindsay Ferrara and Gabrielle Hodge

29 Gesture and Sign: Cataclysmic Break or Dynamic Relations?

Cornelia Müller

49 The Body as Evidence for the Nature of Language

Wendy Sandler

70 Constructing Complexity in a Young Sign Language

Svetlana Dachkovsky, Rose Stamp and Wendy Sandler

91 Different Approaches to Meaning in Primate Gestural and Vocal Communication

Katja Liebal and Linda Oña

2 REPRESENTATION, QUANTIFICATION, AND MODELLING OF MEANINGFUL ELEMENTS

98 Visuo-Kinetic Signs are Inherently Metonymic: How Embodied Metonymy Motivates Forms, Functions, and Schematic Patterns in Gesture

Irene Mittelberg

116 The Palm-Up Puzzle: Meanings and Origins of a Widespread Form in Gesture and Sign

Kensy Cooperrider, Natasha Abner and Susan Goldin-Meadow

132 Iconicity in Signed and Spoken Vocabulary: A Comparison Between American Sign Language, British Sign Language, English, and Spanish

Marcus Perlman, Hannah Little, Bill Thompson and Robin L. Thompson

148 Creating Images With the Stroke of a Hand: Depiction of Size and Shape in Sign Language

Jenny C. Lu and Susan Goldin-Meadow

163 Metaphor in Sign Languages

Irit Meir and Ariel Cohen

176 Visual Iconicity Across Sign Languages: Large-Scale Automated Video Analysis of Iconic Articulators and Locations

Robert Östling, Carl Börstell and Servane Courtaux

193 Automating the Production of Communicative Gestures in Embodied Characters

Brian Ravenet, Catherine Pelachaud, Chloé Clavel and Stacy Marsella

3 PROSODIC STRUCTURE

205 *The Prosodic Characteristics of Non-referential Co-speech Gestures in a Sample of Academic-Lecture-Style Speech*

Stefanie Shattuck-Hufnagel and Ada Ren

218 *Prosody in the Auditory and Visual Domains: A Developmental Perspective*

Núria Esteve-Gibert and Bahia Guellai

228 *Production and Comprehension of Prosodic Markers in Sign Language Imperatives*

Diane Brentari, Joshua Falk, Anastasia Giannakidou, Annika Herrmann, Elisabeth Volk and Markus Steinbach

4 ACQUISITION AND DEVELOPMENT

242 *Using the Hands to Represent Objects in Space: Gesture as a Substrate for Signed Language Acquisition*

Vikki Janke and Chloë R. Marshall

255 *Learning an Embodied Visual Language: Four Imitation Strategies Available to Sign Learners*

Aaron Shield and Richard P. Meier

273 *When Speech Stops, Gesture Stops: Evidence From Developmental and Crosslinguistic Comparisons*

Maria Graziano and Marianne Gullberg



Editorial: Visual Language

Wendy Sandler^{1*}, Marianne Gullberg^{2*} and Carol Padden^{3*}

¹ Sign Language Research Lab, University of Haifa, Haifa, Israel, ² Centre for Languages and Literature, Lund University, Lund, Sweden, ³ Department of Communication and Center for Research in Language, University of California San Diego, San Diego, CA, United States

Keywords: sign language, gesture studies, multimodality, iconicity, visual language

Editorial on the Research Topic

Visual Language

Traditionally, research on human language has taken speech and written language as the main domains of investigation. Visual aspects of language have therefore long been excluded from study. However, the advent of technology allowing us to capture and include auditory and visual signals in the study of language has changed the landscape. There is now a wealth of empirical studies documenting visual aspects of language, ranging from rich studies of sign languages, the most highly sophisticated and self-contained visual language systems, to the burgeoning field of gesture studies, which targets speech-associated gestures, facial expressions, and other bodily movements related to communicative expressions as new domains of study.

However, despite the large body of work now available documenting visual elements of language, sign language and gesture are rarely treated together in theoretical discussions of the human language faculty. Sign language studies often search for linguistic structures that are derived from spoken language theory. Conversely, gesture researchers refrain from defining gestures as “linguistic” (although they often insist that they are part of “language”), because they do not conform to certain properties that linguists consider defining properties, such as strict compositional structure and syntactic rules. In both cases, definitions and concomitant exclusions are not necessarily enlightening, since both domains—speech-associated gestures and sign language—naturally exploit visual expression, and must both be considered in attempting to arrive at a comprehensive account of the human language faculty. By considering both types of visual language, the 19 papers in this *Frontiers Research Topic* volume thus transcend theoretical—and, we would say, artificial—divides. The collection aims to pave the way for an inherently multimodal view of language, in which visible actions of the body play a crucial role.

The volume treats four broad topics: (1) the multimodal nature of language; (2) multi modal representation of meaning; (3) multimodal and multi channel prosody; and (4) acquisition and development of visual language in children and adults. This division aims to organize the Research Topic for the reader, although there is some inevitable overlap.

The first topic targets the nature of all language as multimodal, examining the relationships between speech, gestures, and sign. Visible parts of the body can be engaged in language use in a range of ways, and the papers in this section illustrate specific language phenomena that are multimodal. Perniss; Ferrara and Hodge both review evidence to support a multimodal model of language that accounts for how humans coordinate their semiotic repertoires in crossmodal and composite ways. These authors draw on fundamental modes of communication, including depiction, description, and indicating (Clark, 1996, 2016). Both papers also stress the need to consider the wider context in which utterances are constructed and interpreted, in order to fully understand how multimodal resources are integrated into language as traditionally defined. Müller delves deep into the theoretical debates concerning the status of gestures relative to speech, and addresses the question—are gestures part of language or *are* they language themselves? She further

OPEN ACCESS

Edited and reviewed by:

Manuel Carreiras,
Basque Center on Cognition, Brain
and Language, Spain

*Correspondence:

Wendy Sandler
wendy.sandler@gmail.com
Marianne Gullberg
marianne.gullberg@ling.lu.se
Carol Padden
cpadden@ucsd.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 04 July 2019

Accepted: 15 July 2019

Published: 02 August 2019

Citation:

Sandler W, Gullberg M and Padden C
(2019) Editorial: Visual Language.
Front. Psychol. 10:1765.
doi: 10.3389/fpsyg.2019.01765

discusses the relationship between the speech-gesture ensemble and sign language, specifically targeting the issue of whether the systems are fundamentally different in nature, or whether there is a continuum. Sandler argues for the centrality of the body in understanding the nature of a central property of language: compositionality. She details the linguistic functions of different bodily articulations in the prosodic, lexical, and pragmatic structure of established sign languages, and their recruitment in the emergence of new sign languages, illuminating more general principles of compositionality common to spoken and visual languages alike. The paper goes on to seek possible evolutionary roots of communicative compositionality in physical displays of intense emotions by athletes, and their interpretation. Dachkovsky et al. focus on the relationship between linguistic complexity and its expression by the body, in the emergence of a young sign language, Israeli Sign Language. Drawing on narratives produced by three generations of signers, the authors illustrate how the self-organization of bodily articulations becomes more systematic and reduced as the language becomes more complex over time. Finally, Liebal and Oña discuss the search for the roots of human language in a cross-species comparative approach, and investigate whether precursors to language may already be present in our closest relatives, the non-human primates. They review the debate concerning whether non-human primates use gestures to “mean” the same as humans, and present an overview of how different approaches to visual/gestural vs. vocal communication in non-human primates lead to different answers.

While the first topic deals with different kinds of structure conveyed in language, the second broad topic concerns how meaning can be represented multimodally, and the ways in which meaningful elements can be quantified and modeled. The papers in this section address issues such as how the body, and specifically the hands, can create meaning visually and kinetically in speech-associated gestures and sign languages. Mittelberg begins with a discussion of meaning-making in speech-associated gestures which involves iconicity (a direct form-meaning correspondence), indexicality (contiguity), and habit (conventionality). Comparing two ways in which meaning can be extended in language, metonymy and metaphor, she argues that metonymy is a more basic principle in gestures and signs than metaphor. Mittelberg describes metonymy as more experientially grounded than metaphor, as it highlights a partial aspect of a larger context of human activity, the activity itself being expressed within a *frame*, or a context of experience. Metonymic gestures are simultaneously indexical and refer to conventions of human practice. Cooperrider et al. explore a single gestural form, the so-called epistemic palm up, as a starting point for examining a network of meanings that appear to be similar across gesture and sign. These comparisons serve as the basis for a discussion of the origins of communicative forms, how they divide into multiple different meanings, and become integrated into language. In an unusual comparative study across language modalities, Perlman et al. examine the presence of iconicity in two signed languages (American Sign Language and British Sign Language) and two spoken languages (Spanish and English). The analyses reveal characteristic patterns of iconicity across semantic

domains both within and across the languages depending on the affordances of the main modality. Three further papers focus on iconicity in sign languages specifically: Lu and Goldin-Meadow examine depiction in American Sign Language to reveal a conventional (more lexicalized) and, at the same time, a so-called embellished (more gesture-like) kind of depiction, explaining that the preference depends on context and task. Meir and Cohen investigate metaphors in Israeli Sign Language. They provide a detailed analysis of the ways in which metaphors in sign language differ from metaphors in spoken language, and suggest two principles to account for these differences. They conclude that all human languages exploit metaphorical expression to convey vivid sensory images, while the visual and the auditory modalities impose different constraints on such expression. The fact that the body is visible while signing determines the ways in which signers can refer metaphorically to the body for both human and non-human properties. Finally, in a methodologically oriented paper, Östling et al. use computer based tools to automatically process 120,000 videos from 31 sign languages to reveal two different cross-linguistic patterns of iconicity: the use of two hands to represent plurality, and of locations on different parts of the body to represent activities associated with such locations (e.g., the head with thinking). Computational modeling is a revealing tool for simulating natural communication and testing its interpretation. Ravenet et al. describe the challenges involved in modeling multimodal behavior for so-called Embodied Conversational Agents (ECAs). They identify elements that need to be captured regarding speech and gesture in order to automatically generate multimodal communicative behavior in successful virtual/robotic conversational partners.

The third topic in the volume is concerned with prosody that is multimodal (speech and gesture) and multi channel (manual and non-manual in sign language). Prosody refers to linguistic cues such as intonation, tone, stress, and rhythm, which are superimposed on the morphosyntactic language stream. Both in the domain of sign language and in gesture studies, empirical studies of the coordination of visual prosodic cues with the phrases and sentences of language are quite rare (but for pioneering work, see e.g., Nespor and Sandler, 1999; Sandler, 2010 and Sandler this volume for sign language; McClave, 1994 for gesture). Shattuck-Hufnagel and Ren examine the precise nature of the temporal relationship between speech and one type of co-speech gesture in adults, looking at how non-referential gestures in academic lectures coordinate with prosodic prominence in speech. The analyses reveal a tight link between the prosodic structure of spoken utterances and bodily movements, supporting the claim that a comprehensive speech production model must generate and align gesture and speech as part of the same system. Esteve-Gibbert and Guellai contextualize and evaluate a range of studies on the development of the prosodic coordination of speech and gesture in childhood. Brentari et al. focus on visible prosodic markers in the manual and non-manual channels of different types of imperatives in American Sign Language (ASL). They also test their comprehension by signers of ASL, as well as by signers of a different sign language (German Sign Language, DGS), and finally by hearing non-signers. Results show that different speech

acts display different patterns, and also, importantly, that the patterns are sign language-specific.

The fourth topic deals with acquisition and development of visual language, both in child and adult language learners considering both sign language and gestures. Janke and Marshall ask whether speech-associated gestures function as a useful starting point or scaffold for hearing adults learning sign language, and whether iconic signs are easier to learn than less iconic signs, as is often claimed. The results suggest that adult hearing learners cannot straightforwardly draw on gestures, whether iconic or not. Instead, the challenge seems to be to reduce gestural resources and “linguisticize” a small number of hand shapes to arrive at forms that are part of the grammar of a sign language. Shield and Meier also examine children’s and adults’ acquisition of sign language and posit four possible strategies for learning signs/imitating gestures. They review evidence from typical and atypical hearing and deaf groups to reveal different developmental trajectories across typical and atypical populations. Finally, Graziano and Gullberg focus on the well-rooted assumption that gesture is mainly a compensatory device to support speaking difficulties. Analyses of fluent and disfluent speech from both adult competent speakers of different languages and child and adult language learners instead suggest that gestures are integrated with speech such that both modalities are affected by speech production difficulties. The results thus

support an integrated view of speech and gesture and of a view of language use as fundamentally multimodal.

In conclusion, the papers in this volume provide new evidence for the role of visual elements expressed by the body in language. The volume unifies theoretical and empirical proposals toward a more comprehensive view of the multimodal nature of language, in which speech, gestures, and sign are treated on a par. We hope that the volume will provide additional substance to Perniss’ conclusion that “[W]e are already on the threshold of a new paradigm.”

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

The conceptualization of this volume arose from research project 340140 funded by the European Research Council and led by WS, called The Grammar of the Body (<http://gramby.haifa.ac.il/>). We also acknowledge funding to MG from the Wallenberg Foundations toward her Wallenberg Scholar grant Embodied Bilingualism (MAW 2017.0116).

REFERENCES

- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 3, 324–347. doi: 10.1037/rev0000026
- McClave, E. (1994). Gestural beats: the rhythm hypothesis. *J. Psycholinguist. Res.* 23, 45–66.
- Nespor, M., and Sandler, W. (1999). Prosody in Israeli Sign Language. *Lang. Speech* 42, 143–176. doi: 10.1177/00238309990420020201
- Sandler, W. (2010). Prosody and syntax in sign language. *Trans. Philol. Soc.* 108, 298–328. doi: 10.1111/j.1467-968X.2010.01242.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Sandler, Gullberg and Padden. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Why We Should Study Multimodal Language

Pamela Perniss*

School of Humanities, University of Brighton, Brighton, United Kingdom

Keywords: multimodality, language, communication, sign language, gesture

OPEN ACCESS

Edited by:

Wendy Sandler,
University of Haifa, Israel

Reviewed by:

Mark Aronoff,
Stony Brook University, United States
Jonathan Ginzburg,
Paris Diderot University, France

*Correspondence:

Pamela Perniss
p.perniss@brighton.ac.uk

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 15 December 2017

Accepted: 11 June 2018

Published: 28 June 2018

Citation:

Perniss P (2018) Why We Should
Study Multimodal Language.
Front. Psychol. 9:1109.
doi: 10.3389/fpsyg.2018.01109

What do we study when we study language? Our theories of language, and particularly our theories of the cognitive and neural underpinnings of language, have developed primarily from the investigation of spoken language. Moreover, spoken language has been studied primarily as a unichannel phenomenon, i.e., as just speech or text. However, contexts of face-to-face interaction form the primary ecological niche of language, both spoken and signed, and the primary contexts in which language is used, is learned and has evolved (Levinson and Enfield, 2006; Vigliocco et al., 2014). In such contexts, a multitude of cues, both vocal and visual, contribute to utterance construction. Should we thus not turn our attention to the study of language as multimodal language? The position that language can be appropriately studied as just speech or text essentially aligns with a conception of language based on Chomsky's *competence* or Saussure's *langue*: it is the linguistic code and the formalization of phonological, morphological, and syntactic structure that is of interest. Even functional, usage-based theories of language, which see linguistic structure as shaped by language use and the function of language in cultural and communicative contexts (e.g., Fillmore, 1982; Givón, 1984; Goldberg, 1995), have focused on the linguistic code and have thus also mainly regarded language as speech or text (but see e.g., Tomasello, 1999; Diessel, 2006). The argument put forward here is that we should study language as its multimodal manifestation in contexts of face-to-face interaction. As such, our object of study should subsume information expressed in both vocal and visual channels, including prosody, gesture, facial expression, body movement, which invariably accompany linguistic expression in face-to-face contexts.

The thought experiment proposed by Vigliocco et al. (2014) offers a window onto this approach by asking: *What if the study of language had started with the study of signed language rather than spoken language?* If the study of language had started with signed language, the multichannel/multimodal nature of language would have stood center stage from the beginning. Questions that have become matters of serious inquiry and debate only recently, in particular concerning the status and interplay of iconicity and arbitrariness (Perniss et al., 2010; Perniss and Vigliocco, 2014; Dingemanse et al., 2015) and gradience and categoricity (see Goldin-Meadow and Brentari, 2017 and peer commentary, e.g., Occhino and Wilcox, 2017, for review) in language, may have been discussed earlier and answered in different ways. This brings to the fore the relevance of thinking about language in a more unified way: encompassing spoken and signed language; considering multiple channels of expression; and conceptualizing language with respect to its communicative functions.

What have been considered to be non-linguistic aspects of communication—including gesture, facial expression, body movement—have largely been studied separately from language proper. Multimodality studies, for example, are often framed as offering analyses of social interaction, studying something that is around language, but not studying language as such (see Mondada, 2016 for an overview). Pioneering scholars in the field of gesture studies have long advocated for a conception of gesture that is part and parcel of language (McNeill, 1985, 1992; Kendon, 2004). Nevertheless, this conception has not been adopted on a large scale. In advocating

for a multimodal conception of “language,” it is important to bear in mind the extent to which our objects of study are constructed by an interplay of the present state of theory, technology and discourse (Kuhn, 1962; Foucault, 1972). This point is made by McNeill (1985: 350) when he writes that the division between speech and gesture (or “body language”) is “a cultural artifact, an arbitrary limitation derived from a particular historical evolution”—they are studied separately, though McNeill considers them to be “parts of a single psychological structure.” The conception that “language” is that which is linguistic, while communication is something different—essentially, the Saussurean and Chomskyan heritage—is not given by necessity. As such, it is time to reconceptualize our object of study and to usher in a new paradigm of language theory, a paradigm that focuses on multimodal language, that aligns with the real world use of language and focuses on *doing language* (Andresen, 2014; Kendon, 2014).

The study of sign language and gesture, as communicative expression in the visual modality, has been instrumental in widening the lens of investigation regarding the question of our object of study when we study language. Signed language highlights the fundamental multimodality and semiotic diversity of language. Moreover, the study of sign language, and its comparisons with speech and/or gesture, has highlighted the difficulties of maintaining a principled distinction between the linguistic and non-linguistic, and shown the need for developing analyses that admit a combination of categorical (considered linguistic) and gradient (considered non-linguistic) aspects of language (Liddell, 2003; Johnston, 2013; Kendon, 2014; Vigliocco et al., 2014; Goldin-Meadow and Brentari, 2017). Similarly, gesture and multimodality research has shown that, like signers, speakers make use of a wide range of semiotic resources, combining vocal and visible action in meaning making and utterance construction (e.g., Kendon, 2004; Mondada, 2016). The study of sign and gesture expose our current models of language as too narrowly conceived. The new paradigm for the study of language must acknowledge a range of semiotic practices (exhibiting iconicity, arbitrariness, gradience, categoricity) as fundamental to and constitutive of communicative expression. Below, I outline developments in contemporary research that further attest to the need for incorporating multimodality into our theories of language.

The neuroscientific investigation of language processing is one area in which the distinction between “language” and “communication,” and between “linguistic” and “non-linguistic” elements has been undermined. Recent research has been unable to find strong evidence supporting this distinction in language use. In addition, there is evidence that the brain does not privilege linguistic information in processing. Rather all kinds of context, including multimodal cues, are processed simultaneously and immediately (Hagoort and van Berkum, 2007). Numerous studies have provided evidence for similar processing of gesture and speech in terms of semantic and temporal integration (Özyürek et al., 2007; Hubbard et al., 2009; Straube et al., 2009; Habets et al., 2011; Dick et al., 2014; Yang et al., 2015; Peeters et al., 2017), as well as in terms of perceiving conventionalized meaning (Andric et al., 2013; Wolf et al., 2017). In addition, there is evidence that

prosodic information from visual and vocal channels is treated similarly by the brain, with gestural beats functioning as visual prosody complementary to speech prosody (Biau et al., 2016). Studies also suggest that the use of different cues from context, including co-speech gesture (Skipper, 2014; Weisberg et al., 2017) and visible mouth movements (van Wassenhove et al., 2005), may speed up processing, aiding interpretation through improved prediction, and requiring less allocation of neural resources and thus conserving metabolic resources. Similar processing of semantically meaningful information, regardless of the modality of presentation has, crucially, also been shown for processing of signed and spoken language (MacSweeney et al., 2004) as well as for integration of pictures with sentence context (Willems et al., 2008). Thus, recent evidence from neuroimaging studies does not support a principled divide between linguistic and non-linguistic elements as the legacy of studying language as *competence* or *langue* presupposes. Instead, the evidence suggests that the brain is specially attuned to *doing language* or *linguaging* (Andresen, 2014; Kendon, 2014).

Additional evidence supporting a multimodal view of language comes from recent research that suggests that what has traditionally been considered to be non-linguistic may in fact be subsumable under grammar and susceptible of grammatical description. Floyd (2016), describing the obligatory incorporation of celestial pointing gestures for time-of-day reference, discusses the possibility of modality hybrid grammars, which would incorporate gestural forms into the grammar. Recent work by Schlenker and Chemla (2017), aims to provide evidence for the grammar-like nature of gestures. Similarly, Ginzburg and Poesio (2016) offer a formalization of intrinsically interactional aspects of language, including gestures as well as disfluencies and non-sentential utterances, with the goal of demonstrating their grammatical, rule-governed behavior. This resonates with work by gesture researchers who have sought to define multimodal approaches to grammar (e.g., Mittelberg, 2006; Fricke, 2012), and who have studied aspects of conventionality in gesture, identifying varying degrees of conventionality in form-meaning pairings in gesture, used consistently across speakers within language communities for conveying certain meanings (e.g., Kendon, 1995, 2004; Calbris, 2011; Bressemer and Müller, 2017; Bressemer et al., 2017; Müller, 2017). Similarly, elements in the vocal modality not traditionally considered to be linguistic have been found to exhibit systematic behavior in terms of discursive and interactional function, e.g., research on the use of clicks and percussives (Wright, 2011; Ogden, 2013) and “filled pauses” like *uh* and *um* (Clark and Fox Tree, 2002).

Technological advances in experimental paradigms, data collection and analysis further motivate the need for a new paradigm in the study of language. The need for experimental control has meant that ecological validity, and the study of language in more real-world settings, has often been sacrificed (Hasson and Honey, 2012). Experimental limitations in the past have thus constrained researchers to the study of certain aspects of language. These aspects have happened to align with a *langue/competence*-type object of study, best represented as individual words (spoken or written lexemes) and combinations

of words (spoken or written sentences). “Non-linguistic” elements, e.g., gradient and iconic elements which naturally occur in parallel and simultaneously with the abstractable, formal linguistic elements, were excluded from study (Tromp et al., 2017). In addition, the wider so-called extra-linguistic context, given by the environment—full of visual and acoustic cues—in which language typically occurs was likewise excluded from study (Knoeferle, 2015). However, new methodologies, and in particular, combinations of methodologies (e.g., Virtual Reality environments with ERP, Tromp et al., 2017; eye-tracking with ERP, Knoeferle, 2015) can improve the interpretation of data from a single methodology. Overall, the development of these technologies will support the construction of multimodal language (in the active sense of *doing language*) as the new object of study, which more resembles real-world use of language, rather than being restricted to just one aspect of it (Kendon, 2009). These technologies will allow investigation of the use and processing of language in more ecologically valid, contextually rich and communicatively real-world settings.

Renewed interest in the evolutionary origins of language also points toward a focus on the multimodality of language. One question that has dominated the discourse on theories of language evolution concerns the modality of early communication. Adherents to the “gesture-first” theory of language (e.g., Corballis, 2002, 2017; Tomasello, 2008; Arbib, 2012) claim that symbolic communication originated in the visual-manual modality, and that there was, over time, a transition to the vocal channel as the main carrier of linguistic function. However, eminent gesture researchers like McNeill (1992, 2012) and Kendon (2009, 2017) have claimed that expression in the vocal and visual modalities must have characterized communication from the very start (see also Perlman, 2017). The explanation of a “switch” from the visual to the vocal modality is difficult to motivate, and the tight semantic and temporal orchestration of multiple channels of expression and semiotic resources observable today (from corpus to neuroimaging studies) suggests that utterance construction has always shown this entanglement of modes. In addition, the evidence supporting tight hand-mouth coordination and links between kinesis (e.g., grip) and vocalization (Gentilucci et al., 2001; Kendon, 2009; Vainio et al., 2013) further support a view that gives the “speech-kinesis ensemble” (Kendon, 2009) pride of place in the phylogenetic evolution of language. Interesting perspectives for the interplay of visual and vocal communication supporting language emergence *ab initio* comes from comparative psychology and animal cognition (Leavens, 2003; Gillespie-Lynch et al., 2014) and from the suggestion by Larsson (2015) that the sounds of tool use and locomotion may have contributed to language evolution in a similar way as visible action and motion. Taking “multimodal language” as our object of study would allow a straightforward reconciliation of such findings.

Finally, developments in the fields of multilingualism research and language documentation offer illustrative guides to the changes that need to be generalized in language theory more broadly. The field of multilingualism research has recently been transformed through the notion of *translanguaging*. Researchers

no longer conceive of *code-switching* or even *code-mixing* as an adequate account of the language behavior of bi-/multilingual speakers (Li, 2017). Bi-/multilingual speakers do not switch between or mix different “codes,” as formal systems of language. Rather, they engage in flexible use of diverse semiotic repertoires. Kusters et al. (2017) note that in translanguaging studies, researchers focus on multilingual communication, but without paying attention to multimodal communicative resources; while in multimodality studies, researchers do not attend to multilingual communication. Given the parallels with respect to the focus on a diverse semiotic repertoire and dynamic language practice, Kusters et al. (2017) note the benefits of bringing the fields together, and suggest that the language practices of signers can offer unique insight into the use and negotiation of both multimodal and multilingual repertoires.

Many linguists, especially those studying endangered languages, have adopted practices consistent with the linguistic subdiscipline of language documentation (Himmelmann, 2006). The goal of language documentation goes beyond the production of a (written) grammar of a language. Rather, the goal is documentation of language use and practice in order to create a “lasting, multipurpose record of a language” (Himmelmann, 2006, p. 1). Technological advances have been a boon here as well. Language documentation demands video-recordings of language use on as broad a scale as possible, including different varieties of use, domains of use, and social interaction. This necessarily includes the multimodality of language, and attention to multichannel and semiotically diverse modes of communication. The recognition that the majority of the world is multilingual is also important here, in that it points to the inadequacy of characterizing knowledge of language as residing in an idealized, monolingual speaker in a homogenous language community (Chomsky, 1965). Ansaldo (2010, p. 622) suggests that lessons from monolingual language use and transmission may represent such “exotic communicative ecologies in the history of human language evolution [that] the lessons derived from their study, albeit significant, could well end up being potentially exceptional, maybe even peripheral to the construction of general theories of language.”

Similarly, our models of language need to be based on ecologically valid contexts of multimodal language use (contexts of *doing language*)—and not on the “exotic communicative ecologies” represented by just speech or text. The development of our hitherto dominant models of language has been based on only a part of language, the abstractable, linguistic part best exemplified by written form (McNeill, 1985). A multimodal language model includes the full complement of fundamental modes of communication, including depiction, description, and indexing (Clark, 1996, 2016), and the wider context in which utterances are constructed and interpreted (Kendon, 2014; Vigliocco et al., 2014; Knoeferle, 2015). In various and interconnected ways, the studies reviewed above suggest that we are already on the threshold of a new paradigm. They point to the large range of elements, both vocal and visual, that contribute in systematic ways to language use and communicative expression

and which we should not exclude a priori from the study of language (See Andrén (2014) for discussion of the nature of the problem of delineating the “lower limit of gesture”—the problem of drawing a line between what aspects of “visible action as utterance” Kendon (2004) to include or exclude from study.). We must remind ourselves that science often progresses precisely through a redefinition of the object of study. By redefining the nature and parameters of our concept of “language” we will be capable of forging this new paradigm adequate to a unified conception of language as communication, and basing our theories of language on language as a multimodal phenomenon.

REFERENCES

- Andrén, M. (2014). “On the lower limit of gesture,” in *From Gesture in Conversation to Visible Action as Utterance*, eds M. Seyfeddinipur and M. Gullberg (Amsterdam: John Benjamins), 153–174.
- Andresen, J. (2014). *Linguistics and Evolution: A Developmental Approach*. Cambridge: Cambridge University Press.
- Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., and Small, S. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia* 51, 1619–1629. doi: 10.1016/j.neuropsychologia.2013.03.022
- Ansaldo, U. (2010). Identity alignment and language creation in multilingual communities. *Lang. Sci.* 32, 615–623. doi: 10.1016/j.langsci.2010.08.003
- Arbib, M. (2012). *How the Brain Got Language: The Mirror Neuron Hypothesis*. Oxford: Oxford University Press.
- Biau, E., Moris Fernández, L., Holle, H., Avila, C., and Soto-Faraco, S. (2016). Hand gestures as visual prosody: BOLD responses to audio-visual alignment are modulated by the communicative nature of the stimuli. *Neuroimage* 132, 129–137. doi: 10.1016/j.neuroimage.2016.02.018
- Bressem, J., and Müller, C. (2017). The “Negative-Assessment-Construction” – a multimodal pattern based on a recurrent gesture? *Ling. Vang.* 3, 1–9. doi: 10.1515/lingvan-2016-0053
- Bressem, J., Stein, N., and Wegener, C. (2017). Multimodal language use in Savosavo: refusing, excluding and negating with speech and gesture. *Pragmatics* 2, 173–206. doi: 10.1075/prag.27.2.01bre
- Calbris, G. (2011). *Elements of Meaning in Gesture*. Amsterdam: John Benjamins.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Boston, MA: MIT Press.
- Clark, H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Clark, H. (2016). Depicting as a method of communication. *Psychol. Rev.* 3, 324–347. doi: 10.1037/rev0000026
- Clark, H. H., and Fox Tree, J. E., (2002). Using *uh* and *um* in spontaneous speaking. *Cognition* 84, 73–111. doi: 10.1016/S0010-0277(02)00017-3
- Corballis, M. (2002). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.
- Corballis, M. (2017). “A word in the hand: the gestural origins of language,” in *Neural Mechanisms of Language. Innovations in Cognitive Neuroscience*, ed M. Mody (Boston, MA: Springer), 199–218.
- Dick, A. S., Mok, E. H., Raja Beharelle, A., Goldin-Meadow, S., and Small, S.L., (2014). Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Hum. Brain Mapp.* 35, 900–917. doi: 10.1002/hbm.22222
- Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cogn. Ling.* 17, 463–489. doi: 10.1515/COG.2006.015
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends Cogn. Sci.* 19, 603–615. doi: 10.1016/j.tics.2015.07.013
- Fillmore, C. (1982). “Frame semantics,” in *Cognitive Linguistics: Basic Readings*, ed D. Geeraerts (Berlin: Mouton), 373–400.
- Floyd, S. (2016). Modality hybrid grammar? Celestial pointing for time-of-day reference in Nheengatú. *Language* 92, 31–64. doi: 10.1353/lan.2016.0013
- Foucault, M. (1972). *The Archaeology of Knowledge*. London: Routledge.
- Fricke, E. (2012). *Grammatik multimodal: Wie Wörter und Gesten Zusammenwirken*. Berlin: De Gruyter.
- Gentilucci, M., Benuzzi, F., Gangitano, M., and Grimaldi, S. (2001). Grasp with hand and mouth: a kinematic study on healthy subjects. *J. Neurophys.* 86, 1685–1699. doi: 10.1152/jn.2001.86.4.1685
- Gillespie-Lynch, K., Greenfield, P. M., Lyn, H., and Savage-Rumbaugh, S. (2014). Gestural and symbolic development among apes and humans: support for a multimodal theory of language evolution. *Front. Psychol.* 5:1228. doi: 10.3389/fpsyg.2014.01228
- Ginzburg, J., and Poesio, M. (2016). Grammar is a system that characterizes talk in interaction. *Front. Psychol.* 7:1938. doi: 10.3389/fpsyg.2016.01938
- Givón, T. (1984). *Syntax: A Functional-Typological Introduction*. Vol. I. Amsterdam: John Benjamins.
- Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: Chicago University Press.
- Goldin-Meadow, S., and Brentari, D. (2017). Gesture, sign, and language: the coming of age of sign language and gesture studies. *Behav. Brain Sci.* 40:e46. doi: 10.1017/S0140525X15001247
- Habets, B., Kita, S., Shao, Z., Özyürek, A., and Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *J. Cogn. Neurosci.* 23, 1845–1854. doi: 10.1162/jocn.2010.21462
- Hagoort, P., and van Berkum, J. (2007). Beyond the sentence given. *Philos. Trans. R. Soc. B* 362, 801–811. doi: 10.1098/rstb.2007.2089
- Hasson, U., and Honey, C. (2012). Future trends in neuroimaging: neural processes as expressed within real-life contexts. *Neuroimage* 62, 1272–1278. doi: 10.1016/j.neuroimage.2012.02.004
- Himmelman, N. (2006). “Language documentation: what is it and what is it good for?” in *Essentials of Language Documentation*, eds J. Gippert, N. Himmelman, and U. Mosel (Berlin: Mouton), 1–30.
- Hubbard, A., Wilson, S., Callan, D., and Dapretto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Hum. Brain Mapp.* 30, 1028–1037. doi: 10.1002/hbm.20565
- Johnston, T. (2013). Towards a comparative semiotics of pointing actions in signed and spoken languages. *Gesture* 13, 109–142. doi: 10.1075/gest.13.2.01joh
- Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *J. Pragm.* 23, 247–279. doi: 10.1016/0378-2166(94)00037-F
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kendon, A. (2009). Language’s matrix. *Gesture* 9, 355–372. doi: 10.1075/gest.9.3.05ken
- Kendon, A. (2014). Semiotic diversity in utterance production and the concept of ‘language’. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130293. doi: 10.1098/rstb.2013.0293
- Kendon, A. (2017). Reflections on the “gesture-first” hypothesis of language origins. *Psychon. Bull. Rev.* 24, 163–170. doi: 10.3758/s13423-016-1117-3
- Knoefler, P. (2015). “Language comprehension in rich non-linguistic contexts: combining eye tracking and event-related brain potentials,” in *Cognitive*

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

ACKNOWLEDGMENTS

I thank the editor and reviewers for helpful comments on an earlier version of the article. I thank the School of Humanities, University of Brighton for providing the funds to cover open access publishing fees.

- Neuroscience of Natural Language Use, ed R. M. Willems (Cambridge: Cambridge University Press), 77–100.
- Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago, IL: Univ. Chic. Press.
- Kusters, A., Spotti, M., Swanwick, R., and Tapio, E. (2017). Beyond languages, beyond modalities: transforming the study of semiotic repertoires. *Int. J. Multiling.* 14, 219–232. doi: 10.1080/14790718.2017.1321651
- Larsson, M. (2015). Tool-use-associated sound in the evolution of language. *Anim. Cogn.* 18, 993–1005. doi: 10.1007/s10071-015-0885-x
- Leavens, D. (2003). Integration of visual and vocal communication: evidence for Miocene origins. *Behav. Brain Sci.* 26, 232–233. doi: 10.1017/S0140525X03480060
- Levinson, S., and Enfield, N. (2006). *The Roots of Human Sociality: Culture, Cognition and Interaction*. London: Bloomsbury.
- Li, W. (2017). Translanguaging as a practical theory of language. *Appl. Ling.* 39, 9–30. doi: 10.1093/applin/aml039
- Liddell, S. (2003). *Grammar, Gesture and Meaning in American Sign Language*. Cambridge: Cambridge University Press.
- MacSweeney, M., Campbell, R., Woll, B., Giampietro, V., David, A. S., McGuire, P., et al. (2004). Dissociating linguistic and nonlinguistic gestural communication in the brain. *Neuroimage* 22, 1605–1618. doi: 10.1016/j.neuroimage.2004.03.015
- McNeill, D. (1985). So you think gestures are nonverbal? *Psych. Rev.* 92, 350–371. doi: 10.1037/0033-295X.92.3.350
- McNeill, D. (1992). *Hand and Mind*. Chicago, IL: Univ. Chic. Press.
- McNeill, D. (2012). *How Language Began. Gesture and Speech in Human Evolution*. Cambridge: Cambridge University Press.
- Mittelberg, I. (2006). *Metaphor and Metonymy in Language and Gesture: Discourse Evidence for Multimodal Models of Grammar*. Ann Arbor, MI: UMI.
- Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *J. Socioling.* 20, 336–366. doi: 10.1111/josl.1_12177
- Müller, C. (2017). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture* 16, 277–304. doi: 10.1075/gest.16.2.05mul
- Occhino, C., and Wilcox, S. (2017). Gesture or sign? a categorization problem. *Behav. Brain Sci.* 40, 36–37. doi: 10.1017/S0140525X15003015
- Ogden, R. (2013). Clicks and percussives in English conversation. *J. Int. Phonetic Assoc.* 43, 299–320. doi: 10.1017/S0025100313000224
- Özyürek, A., Willems, R., Kita, S., and Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cogn. Neurosci.* 19, 605–616. doi: 10.1162/jocn.2007.19.4.605
- Peeters, D., Snijders, T., Hagoort, P., and Özyürek, A. (2017). Linking language to the visual world: neural correlates of comprehending verbal reference to objects through pointing and visual cues. *Neuropsychologia* 95, 21–29. doi: 10.1016/j.neuropsychologia.2016.12.004
- Perlman, M. (2017). Debunking two myths against vocal origins of language: language is iconic and multimodal to the core. *Interact. Stud.* 18, 376–401. doi: 10.1075/is.18.3.05per
- Perniss, P., and Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philos. Trans. R. Soc. B* 369:20130300. doi: 10.1098/rstb.2013.0300
- Perniss, P., Thompson, R., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol.* 1:227. doi: 10.3389/fpsyg.2010.00227
- Schlenker, P., and Chemla, E. (2017). Gestural agreement. *Nat. Lang. Linguist. Theory* 36, 587–625. doi: 10.1007/s11049-017-9378-8
- Skipper, J. (2014). Echoes of the spoken past: how auditory cortex hears context during speech perception. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130297. doi: 10.1098/rstb.2013.0297
- Straube, B., Green, A., Weis, S., Chatterjee, A., and Kircher, T. (2009). Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance. *J. Cogn. Neurosci.* 21, 821–836. doi: 10.1162/jocn.2009.21053
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.
- Tomasello, M. (2008). *The Origins of Human Communication*. Cambridge: MIT Press.
- Tromp, J., Peeters, D., Meyer, A., and Hagoort, P. (2017). The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behav. Res. Methods* 50, 862–869. doi: 10.3758/s13428-017-0911-9
- Vainio, L., Schulman, M., Tiippana, K., and Vainio, M. (2013). Effect of syllable articulation on precision and power grip performance. *PLoS ONE* 8:e53061. doi: 10.1371/journal.pone.0053061
- van Wassenhove, V., Grant, K., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1181–1186. doi: 10.1073/pnas.0408949102
- Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130292. doi: 10.1098/rstb.2013.0292
- Weisberg, J., Hubbard, A., and Emmorey, K. (2017). Multimodal integration of spontaneously produced representational co-speech gestures: an fMRI study. *Lang. Cogn. Neurosci.* 32, 158–174. doi: 10.1080/23273798.2016.1245426
- Willems, R., Özyürek, A., and Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *J. Cogn. Neurosci.* 20, 1235–1249. doi: 10.1162/jocn.2008.20085
- Wolf, D., Rekkittke, L.-M., Mittelberg, I., Klasen, M., and Mathiak, D. (2017). Perceived conventionality in co-speech gestures involves the fronto-temporal language network. *Front. Hum. Neurosci.* 11:573. doi: 10.3389/fnhum.2017.00573
- Wright, M. (2011). On clicks in English talk-in-interaction. *J. Int. Phonetic Assoc.* 41, 207–229. doi: 10.1017/S0025100311000144
- Yang, J., Andric, M., and Mathew, M. (2015). The neural basis of hand gesture comprehension: a meta-analysis of functional magnetic resonance imaging studies. *Neurosci. Biobehav. Rev.* 57, 88–104. doi: 10.1016/j.neubiorev.2015.08.006

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Perniss. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Language as Description, Indication, and Depiction

Lindsay Ferrara^{1*} and Gabrielle Hodge²

¹ Department of Language and Literature, Norwegian University of Science and Technology, Trondheim, Norway, ² Deafness Cognition and Language Centre, University College London, London, United Kingdom

OPEN ACCESS

Edited by:

Wendy Sandler,
University of Haifa, Israel

Reviewed by:

Pamela Permiss,
University of Brighton,
United Kingdom
Silva Ladewig,
European University Viadrina,
Germany
Jana Bressem,
Technische Universität Chemnitz,
Germany

*Correspondence:

Lindsay Ferrara
lindsay.n.ferrara@ntnu.no

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 12 November 2017

Accepted: 24 April 2018

Published: 23 May 2018

Citation:

Ferrara L and Hodge G (2018)
Language as Description, Indication,
and Depiction. *Front. Psychol.* 9:716.
doi: 10.3389/fpsyg.2018.00716

Signers and speakers coordinate a broad range of intentionally expressive actions within the spatiotemporal context of their face-to-face interactions (Parmentier, 1994; Clark, 1996; Johnston, 1996; Kendon, 2004). Varied semiotic repertoires combine in different ways, the details of which are rooted in the interactions occurring in a specific time and place (Goodwin, 2000; Kusters et al., 2017). However, intense focus in linguistics on conventionalized symbolic form/meaning pairings (especially those which are arbitrary) has obscured the importance of other semiotics in face-to-face communication. A consequence is that the communicative practices resulting from diverse ways of being (e.g., deaf, hearing) are not easily united into a global theoretical framework. Here we promote a theory of language that accounts for how diverse humans coordinate their semiotic repertoires in face-to-face communication, bringing together evidence from anthropology, semiotics, gesture studies and linguistics. Our aim is to facilitate direct comparison of different communicative ecologies. We build on Clark's (1996) theory of language use as 'actioned' via three methods of signaling: describing, indicating, and depicting. Each method is fundamentally different to the other, and they can be used alone or in combination with others during the joint creation of multimodal 'composite utterances' (Enfield, 2009). We argue that a theory of language must be able to account for all three methods of signaling as they manifest within and across composite utterances. From this perspective, language—and not only language use—can be viewed as intentionally communicative action involving the specific range of semiotic resources available in situated human interactions.

Keywords: sign language, multimodal, semiotics, language, indexicality, depiction

INTRODUCTION

How do humans communicate with each other? One might say there are many paths up the mountain: a hearing speaker describes the use of a basket fish trap by closely aligning his speech with manual gestures depicting the shape of the trap and how it functions (Enfield, 2009, p. 188); a deaf signer unifies lexicalized manual signs within a bodily re-enactment of herself as a young child to express the sense of surprise and wonder she experienced as she learned signed language for the first time (Fenlon et al., 2018, p. 96); while a deafblind signer reaches for the hand of a hearing shopkeeper, gestures "how much?", and then invites the shopkeeper to trace numbers on his palm (Kusters, 2017, p. 400). In each context, each individual engages with others in their environment on their own terms, making use of the various bodily articulators (a voice, hands, body) and strategies for communicating (speech, visible and tactile actions, numerical symbols)

available to them in that moment and physical space. In doing so, they position themselves as independent agents embedded within an intricate and dynamic network of social relationships, someone who effects social actions and is affected by others' actions in turn (Levinson and Enfield, 2006; Enfield, 2013).

Despite the sheer variety of communicative practices that can be observed, many linguists have historically been interested in the question of how 'language' – defined as symbolic, conventionalized, and paradigmatic arrangements for making meaning – works. This has typically involved analyzing communicative phenomena using a Saussure-inspired semiological approach in which the linguistic *signe* is viewed as a dual entity of 'signifier' and 'signified.' The focus has therefore been on those symbolic and conventional pairings of form and meaning that are componential (e.g., phonology, morphosyntax) and therefore easier to identify and analyze. Within this paradigm, the arbitrariness of symbolic *signe* relationships and their potentially decontextualized semantic power is emphasized, while the contextual rootedness and emergent meaningfulness of semiosis (namely, indexicality and iconicity) is often omitted (Parmentier, 1994, p. 5). Yet the aspects of language use which can be analyzed from a structuralist perspective are only part of the picture of how we engage in social actions and communicate: they do not explain everything.

While useful for understanding unimodal patterns of language use, such as the constituency-based analysis of speech or writing, these conventional symbol-driven approaches have resulted in theories of language that do not fully consider the semiotic plurality of human communication, nor how this plurality interacts with the emergence of such conventional symbols. Many researchers have challenged this narrow view of language and have shown how multimodal approaches to language description are necessary for a holistic understanding of human communication. For example, researchers from the field of gesture studies have investigated how to classify and analyze different types of co-speech gestures (e.g., McNeill, 1992; Goldin-Meadow, 2003; Streeck, 2009), including the identification of different types of gestures with respect to their function and degrees of conventionalization and grammaticalization (e.g., Kendon, 2004; Wilcox, 2007; Calbris, 2011; for an overview see Müller et al., 2013, 2014, especially Bressem, 2013). Signed language linguists have investigated the coordination of different types of signs and strategies for making meaning used by deaf signers (e.g., Sutton-Spence and Woll, 1999; Liddell, 2003; Johnston, 2012; Vigliocco et al., 2014), including recent efforts to directly compare the communication of deaf signers with hearing co-speech gesture (see Perniss et al., 2015, *inter alia*).

However, there has yet to be a general theory that unifies this evidence to account for diverse communicative practices. Furthermore, many researchers continue to work within paradigms that posit boundaries between 'language' and 'gesture,' 'linguistic' and 'non-linguistic,' 'verbal,' and 'non-verbal' (see Kendon, 2014). However, as Kendon (2014, p. 3) has argued, "we must go beyond the issue of trying to set a boundary between 'language' and 'non-language,' and occupy ourselves, rather, with an approach that seeks to distinguish these different systems, at the same time analyzing their interrelations." How else can we

directly and systematically compare the communicative practices used by the hearing fisherman and his interactant with those used by the deaf signer and her friend, or the deafblind signer and the shopkeeper? If elements of some repertoires are excluded, our understanding of the complex nature of language variation and diversity cannot progress. Our approach is rather to seek an understanding of how diverse humans (e.g., hearing, deaf) communicate using the semiotic repertoires available to them, and how the resulting conventions of these ecologies can be described empirically. To do this, we build upon Clark's (1996) theory of language use as 'actioned' via three methods of signaling: describing, indicating, and depicting. These methods differ fundamentally in how they signify referents, yet each can be used alone or in combination with others during the joint creation of multimodal 'composite utterances' to effect social actions (Enfield, 2009).

We use Clark's (1996) theory as a starting point, because it is based upon the foundational semiotic principles of 'symbols, indices and icons' first proposed by Peirce (1955). While other linguists and gesture researchers have also advocated Peircean-inspired semiotic approaches for analyzing multimodal language data (e.g., Mittelberg, 2008; Fricke, 2014) – and these approaches are certainly complementary to the one described here – we believe that Clark's theory most clearly marries Kendon's call for a 'comparative semiotics' of signed and spoken communication (Kendon, 2008) with existing semiotic approaches adopted by signed and spoken language linguists (e.g., Liddell, 2003; Enfield, 2009; Dingemanse, 2011; Johnston, 2013b). In taking a semiotic approach (rather than a formal linguistic or gesture-oriented one), we also strive for a modality-free understanding of the function and use of different semiotic acts, and therefore avoid issues that have arisen in approaches which do not consider more gradient aspects of meaning (see Okrent, 2002; Liddell, 2003). In the following sections, we review the literature on communicative practices and semiotic repertoires from an ecological perspective (Haugen, 1972; Goodwin, 2000). We describe Clark's three methods of signaling and the notion of the composite utterance (Enfield, 2009). We then bring together evidence from existing signed and spoken language research, and present examples of composite utterances from deaf signers and hearing speakers. All examples are reflective of the everyday practices signers and/or speakers use to describe, indicate, and/or depict meaning during their interactions. Finally, we conclude with some thoughts on re-orienting language theory to account for these varied communicative practices—thereby underscoring that a theory of language use should not be fundamentally different from a theory of language.

COMMUNICATION PRACTICES AND SEMIOTIC REPERTOIRES

The first step in investigating the communication practices of diverse humans is to consider the communicative ecologies in which these practices emerge. Signers and speakers live in richly dynamic communicative ecologies, in which what we understand as 'language' is just one of many resources available

for making meaning (see Bühler, 1990/1934; Parmentier, 1994; Enfield, 2009; Keevallik, 2018). We coordinate varied bodily articulators (a voice, hands, body) and physical artifacts (e.g., paper, sand, mobile phone) to express communicative intent, the details of which are embedded within interactions occurring in a specific time and place. For example, in the Western desert region of Australia, Ngaanyatjarra children may incorporate alphabetic symbols into their stories drawn in the sand, along with the more traditional iconographic drawings and objects used by adults to index and depict referents in these stories. This youth-driven contribution to established sand story practices reflects generational literacy differences (Kral and Ellis, 2008; see also Green, 2014). Shared semiotic resources and modes of communication within ecologies may therefore be used in different ways by different individuals at different times.

In this sense then, a communicative ecology is not simply the environment in which signers and speakers act; it is the constantly emerging complex shape and history of interactions between language users and their environment (Haugen, 1972; Goodwin, 2000). These reciprocal, dynamic interactions give rise to ‘structural couplings’ (Maturana and Varela, 1987) between individuals and their environment, which manifest as varied communication practices. These practices evolve as signers and speakers draw on all meaningful resources available to them into a complete, heteroglossic package, i.e., the “semiotic repertoire” (Kusters et al., 2017). Within this cognitive/biosemiotics approach, a key principle is that the meanings which emerge within ecologies are largely inferential – more so than symbolic – so that tokens of expression stand in relation to each other with respect to their specific indexical properties (Peirce, 1955; see Kravchenko, 2006).

Another, closely-related principle is that the communication practices which emerge are embedded in the physical environment in which they occur (Duranti and Goodwin, 1992; Goodwin, 2000; Keevallik, 2018). This leads to the emergence of “spatial repertoires” which are defined by the communicative resources available to interactants in a particular place (Neville et al., 2014; Pennycook and Otsuji, 2014). Encounters between agents in an ecology are developed and maintained over various time frames, with the effect that “future interactions occur in a new and adaptive way” (Pickering, 1997, p. 192). Small-scale social encounters between individuals shape larger scale practices and vice versa (Agha, 2005, p. 12). Consequently, communicative practices and repertoires share similarities and differences, both within specific interactions and across social networks, depending on where they unfold (Bourdieu, 1991; Agha, 2007; see also Bernstein, 2003/1971).

Diverse semiotic resources and modes of communication are used to disambiguate the situated context, whereby disambiguation is negotiated between interactants during social interactions via ostensive and inferential acts (LaPolla, 2003, 2005). These notions challenge generative understandings of situated context as being used to disambiguate fixed symbolic forms, whereby the interpretation of ostensive-inferential communication involves a coding-decoding process (cf. Sperber and Wilson, 1986; Wilson and Sperber, 1993). However, it is important to note that an individual’s repertoire

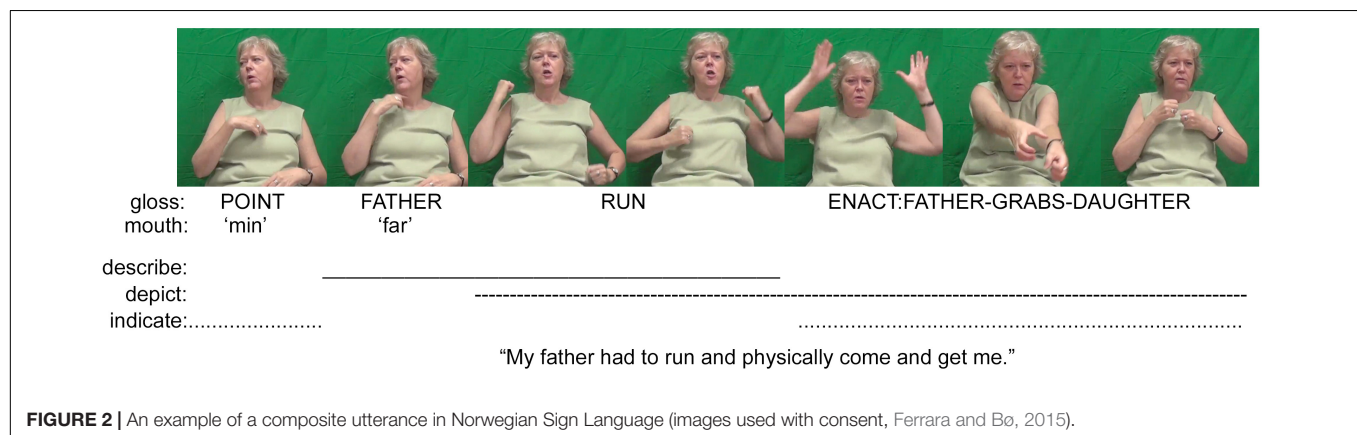
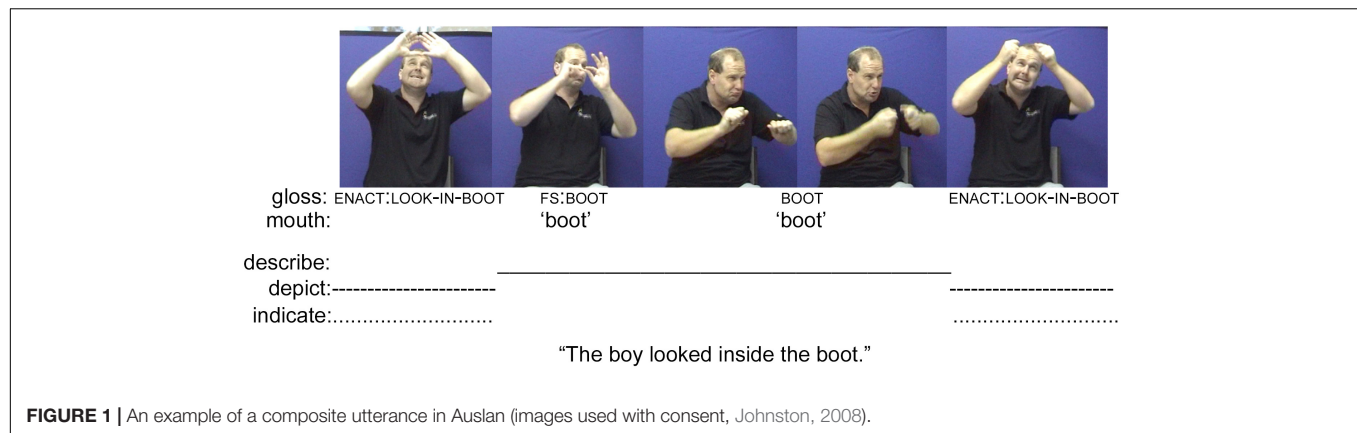
is as much determined by the resources they do not have, as by those they do have (Busch, 2015, p. 14). This factor gains prominence, for example, during interactions between signers and/or speakers whose repertoires do not fully align, as they must actively negotiate which bits of each other’s repertoire can be used effectively – or not (see e.g., Green, 2015; Harrelson, 2017; Hodge, forthcoming). Crucially, an acknowledgment of semiotic diversity enables investigations of signed and spoken languages to relax from the restraints of ‘structure’ and ‘descriptive representation’ resulting from the lineage of de Saussure’s important contributions to linguistics. It re-establishes semiotic diversity as a foundation upon which to identify and explore patterns of embodied communication, of which conventionalized descriptive signaling is just one method, as we will see in the following sections.

P-SIGNS SIGNALLED THROUGH DESCRIPTION, INDICATION, AND DEPICTION

The emergence of diverse communicative practices can be at least partly attributed to the quintessentially face-to-face and multimodal nature of human interactions (Bavelas et al., 1997; Kelly, 2002, 2006; Kita, 2003; Tomasello, 2003; Cienki and Müller, 2008; Calbris, 2011; Müller et al., 2013, 2014). Indeed, the availability of space during face-to-face interactions between deaf signers has been suggested as “a fact that may influence, and even constrain, the linguistic [i.e., communicative] system in other ways” (Johnston, 1996, p. 1). These influences and constraints manifest in the extensive and habitual integration of tokens of three types of signs (in a Peircean sense) in face-to-face, situated discourse: (1) symbols, (2) indices, and (3) icons (Peirce, 1955; see also Parmentier, 1994; Kockelman, 2005; Mittelberg, 2008; Enfield, 2009; Fricke, 2014). Here we refer to tokens of these types of signs as ‘P-signs’ to avoid confusion with other uses of the term ‘sign.’ Clark (1996) proposed that symbols, indices, and icons are signaled through acts of describing, indicating and depicting.¹ Language use is therefore a system of signaling with these three different methods.

Symbols are form-meaning pairings where it is ‘pre-agreed’ that *X* stands for *Y*. Tokens of symbols are fully conventionalized and thus have both token and type identities (Enfield, 2009, p. 13). Examples of symbols include the lexicalized manual signs of deaf signed languages (e.g., the Auslan sign BOOT in **Figure 1** and the Norwegian Sign Language sign FATHER in **Figure 2**), alternate signed languages (see e.g., Kendon, 1988; Green, 2014), as well as the spoken or written words of spoken languages (e.g., the English words *booking a flight* in **Figure 5**). It also includes culturally-specific emblematic manual gestures such as the OK and THUMBS-UP gestures (see e.g., Sherzer, 1991), and even conventionalized intonation contours, such as in the English

¹Clark (1996) first proposed ‘describing-as,’ ‘indicating’ and ‘demonstration’ as the names of the three methods of signaling. Here we abbreviate ‘describing-as’ to ‘describing’ and use the term ‘depiction’ instead of ‘demonstration’ to correspond with more recent signed and spoken language literature (Liddell, 2003; Dudis, 2011; Dingemanse, 2014; Clark, 2016).



utterance "That was cold!" to mean cold-hearted (Liddell, 2003, pp. 358–361), or those whistled by Pirahã men when hunting (Everett, 2005).

Clark (1996) proposes that symbols are signaled through acts of description. It is these descriptions that have been the primary focus of linguistics. Dingemanse (2015) provides an apt characterization of descriptions:

Descriptions are typically arbitrary, without a motivated link between form and meaning. They encode meaning using strings of symbols with conventional significations, as the letters in the word "pipe" or the words in a sentence like "the ball flew over the goal." These symbols are discrete rather than gradient: small differences in form do not correspond to analogical differences in meaning. To interpret descriptions, we decode such strings of symbols according to a system of conventions (Dingemanse, 2015, pp. 950–951).

It is true that understanding how description works is essential to language and linguistic theory. However, it is also true that actual utterances unfolding as parts of specific interactions and spatiotemporal contexts involve much more than description: utterances must index actual referents and meanings, and may therefore also include indices and depictions (Clark, 1996, pp. 161–162).

Indices are forms that anchor communicative events to a specific time and place. These forms are physically connected to their referents, e.g., through finger pointing, and work to

create focused joint attention (Clark, 1996, pp. 164–165). Indices, as opposed to symbols, exhibit both conventional and non-conventional properties. Enfield (2009, p. 13) describes tokens of indices as partly-conventional symbolic indexicals that "[glue] things together, including words, gestures, and (imagined) things in the world." These indexed referents may be physically present and jointly attended, or they may be entirely conceptual and mapped onto a jointly attended real space (Liddell, 1995). Indicating is therefore the method of signaling specific referents via indices using a variety of forms (Clark, 1996). For example, hearing speakers often signal indices using deictic forms such as the English function words *it* and *this*, as well as hand-pointing, lip-pointing, and other culturally-specific bodily actions during which speakers or signers extend parts of their body (or objects that act as an extension of their body) in a direction toward, or contacting, some referent in the context of the utterances (Clark, 1996; Kita, 2003; see also Fricke, 2014). The placement of material objects in a purposeful way in various settings is also a method of indicating (Clark, 2003).

The physical manifestation of pointing actions may also depend on whether agents within a given ecology preference signed or spoken modes of communication, as well as other constraints such as local, culturally-specific conventions and frequency of use. For example, analysis of pointing actions by speakers of Arrernte in Northern Australia has shown that the physical manifestation of these actions is culturally specific

(not universal), with different forms potentially differentiating distinct frames of reference and semantic fields such as near vs. far proximity, absolute vs. relative space, and/or singular vs. plural entities (Wilkins, 2003). Corpus-based analysis of pointing actions produced by deaf native and near-native signers of Auslan (Australian signed language) from a semiotic perspective suggested that pointing actions in signed languages are not fundamentally different to the co-speech pointing actions produced by hearing speakers, and that the linguistic analysis of signed language pointing as fully grammaticalized pronominal forms may not be warranted (Johnston, 2013a,b).

However, one recent comparison of pronominal pointing in the BSL (British Sign Language) Corpus and the Tavis Smiley American English dataset found that the self- and other-directed pointing actions produced by deaf native signers of BSL are more conventionalized and reduced in form compared to those produced by hearing non-signing speakers of American English, although the function of these pointing acts requires further investigation (Fenlon et al., 2016; see also Cormier et al., 2013a; Goldin-Meadow and Brentari, 2015). Within a different language community, the Nheengatú of Brazil, Floyd (2016) found a quite conventional multimodal pattern used to reference time. In this community, speakers produce an auditory articulation coupled with a point to the sun's position to refer to different times of day. Thus regardless of potential fine-grained differences across ecologies, it is evident that both signers and speakers systematically use bodily actions to index physical and abstract referents during their face-to-face interactions. These actions must therefore be included in a theory of language alongside forms that have received more attention from linguists, such as spoken or written deictic markers and pronominal forms, because they are all essential to understanding how humans signal through indicating.

Icons, in contrast to symbols and indices, partially depict meaning through perceptual resemblances (Clark, 1996). Signaling with icons is achieved through 'demonstrations' (Clark, 1996) or 'depictions' (Liddell, 2003). Paintings and drawings are prototypical examples of exhibited depictions, but here we focus on performed depictions co-created between signers and/or speakers (see Clark, 2016). More specifically, depictions are:

[T]ypically iconic, representing what they stand for in terms of structural resemblances between form and meaning. They use material gradiently so that certain changes in form imply analogical differences in meaning. Consider the varying intensity of the strokes of paint that represent the shimmer and shadows on Magritte's pipe, or the continuous movement of a hand gesture mimicking the trajectory of a ball. To interpret depictions, we imagine what it is like to see the thing depicted (Dingemanse, 2015, p. 950).

Depiction signals icons that vary in their degree of conventionalization across a community. For instance, mimetic bodily enactments of people, animals or things (also known as 'constructed action' and 'constructed dialog,' Tannen, 1989; Metzger, 1995) used by signers and speakers to 'show' meaning rather than describe it (see Cormier et al., 2015b) are often analyzed as 'singular events' during which interactants interpret

a form as 'standing for' a meaning within a specific usage event (Kockelman, 2005). These standing-for relations "become signs only when taken as signs in context" (Enfield, 2009, p. 13) (see the enactment by the Auslan signer in **Figure 1** as well as the constructed dialog produced by the English speaker in **Figure 5**).

Across the world's signed languages, signs often called either 'depicting' signs (analyzed as partly lexical signs composed of conventional and non-conventional elements, see Liddell, 2003) or 'classifier' signs (analyzed as signed manifestations of the spoken or written classifier morphemes used in many spoken languages, see Supalla, 2003) represent another way signers can depict meanings. These signs have been a major focus of signed language research and describing and accounting for them within formal and structural theories of language presented an early challenge for signed language linguists (see e.g., Supalla, 1978; Klima and Bellugi, 1979), while others emphasized the iconic and context-dependent nature of these signs (e.g., DeMatteo, 1977; Johnston, 1989; Cogill-Koez, 2000). Researchers have observed that depicting signs are both iconically and spatially motivated while also exhibiting some level of conventionalization. They function to depict the handling of entities, the size and shape of entities, the location of entities, and the movement of entities (e.g., Liddell, 2003; Johnston and Schembri, 2007).

Depicting signs have been compared in varying degrees to the iconic and metaphoric manual gestures (also known as referential gestures) produced as part of spoken language discourse (see e.g., cf. Emmorey, 2003; Schembri et al., 2005; Streeck, 2009; Cormier et al., 2012). In addition, researchers investigating co-speech gesture have established fine-grained methods for detailing how hearing speakers depict with their hands and prompt meaning construction through different types of iconicity—often making a distinction between the hands as they depict the hands doing various activities vs. the hands depicting another type of referent (Müller, 1998, 2014, 2016; Kendon, 2004; Streeck, 2009). The types of gesture that result from these 'modes of gestural representation' are observed to align with the manual enactments and depicting signs observed across signed languages (Streeck, 2009; Müller, 2014).

The manual depictions briefly detailed above can be compared with ideophones. Dingemanse (2011, 2014, 2017a) explains that ideophones are spoken words that depict sensory imagery, and which are more or less integrated with surrounding morphosyntax. Examples include the Japanese *gorogoro* "rolling" and *kibikibi* "energetic" (mentioned in Dingemanse, 2017b). Ideophones function dually as descriptions and depictions, because of their conventionalized status, although novel ideophones can also be created within the context of an interaction. Others have compared ideophones to iconic, lexical signs in signed languages (e.g., Bergman and Dahl, 1994; Ferrara and Halvorsen, 2017). In "Composite Utterances Evidenced Within Hearing/Hearing interactions," we will present an example from a Siwu language interaction that includes two examples of ideophones to illustrate the multimodal, composite utterances produced by hearing speakers.

Before discussing how *P*-signs are coordinated during face-to-face interaction, it is important to note that symbols, indices


and icons are not exclusive categories—as illustrated by the introduction to ideophones above. Following Peirce, Clark (1996, p. 159) notes that “a single sign may have iconic, indexical, and symbolic properties” (emphasis in the original). For example, instances of enactment in which a speaker re-constructs an earlier dialog of themselves or another person might primarily be interpreted as depictions, but they are more precisely *depictions of prior acts of description*. Each depiction (via enactment) of the earlier event indexes both the original act of describing and any subsequent depiction of this event. Ideophones are fully conventional words that have both symbolic and iconic properties (Dingemanse, 2011). Signed language *P*-signs also exhibit multiple properties. Fully conventional lexical signs are descriptions, but in the case of more iconic lexical signs, they can also be used as depictions (e.g., the token of the lexical sign RUN in **Figure 2**, see also Johnston and Ferrara, 2012; Ferrara and Halvorsen, 2017). Other signs can be both symbolic and indexical, such as fully lexical signs that are meaningfully directed in space to index a referent (Liddell, 2003; Cormier et al., 2015a).

COMPOSITE UTTERANCES IN SIGNED AND SPOKEN LANGUAGES

Signers and speakers combine the three types of *P*-signs to ‘tell, show and do’ during face-to-face interactions. This occurs via the mutual orientation, recognition, and interpretation of social acts defined as communicative ‘moves.’ Within communicative moves, tokens of *P*-signs are temporally and spatially coordinated to create unified ‘composite utterances’ that are interpreted holistically rather than componentially

(Enfield, 2009). A communicative move may be recognized as part of an interactional sequence, such as a turn, or more specifically as an instantiation of a type of linguistic utterance, such as an intonation unit or clause (see e.g., Thompson and Couper-Kuhlen, 2005). These moves are further defined by the temporal domain of ‘conversation time,’ i.e., the moment-by-moment temporality in which communicative moves unfold. Enfield (2009, p. 10) uses the term ‘enchrony’ to refer to conversation time and to differentiate it from historical time, i.e., diachrony.


As products of face-to-face interactions, composite utterances can be analyzed according to both their semiotic properties and the situated context of the interactions in which they emerge. With respect to their interpretation, it is the interaction of the elements within the composition that drives the creation – or rather, the disambiguation – of a “precise and vivid understanding” (Kendon, 2004, p. 174) more so than the use of language *per se* (see also Armstrong et al., 1995). The preciseness and vividness of an understanding, however, might be clarified by using more conventionalized semiotic resources such as lexicalized words or signs, to frame the less conventionalized properties of the utterance. For example, deaf signers’ strategic use of lexicalized signs to index and frame subsequent token enactments work to clarify who or what is being vividly enacted. In the same way, it is often the case that the visible bodily actions created by hearing speakers “cannot be precisely interpreted until [they are] perceived as part of the gesture-speech ensemble in which [they are] employed” (Kendon, 2004, p. 169). However, this relationship is reciprocal. For example, a hearing speaker’s enactment of throwing rice on the ground makes more salient the more vivid aspects of the verb ‘throw’ uttered in the speech, while



gloss: SAME DS:TABLE-SMALL DS:TABLE-MEDIUM DS:TABLE-LARGE

describe: _____

depict: -----



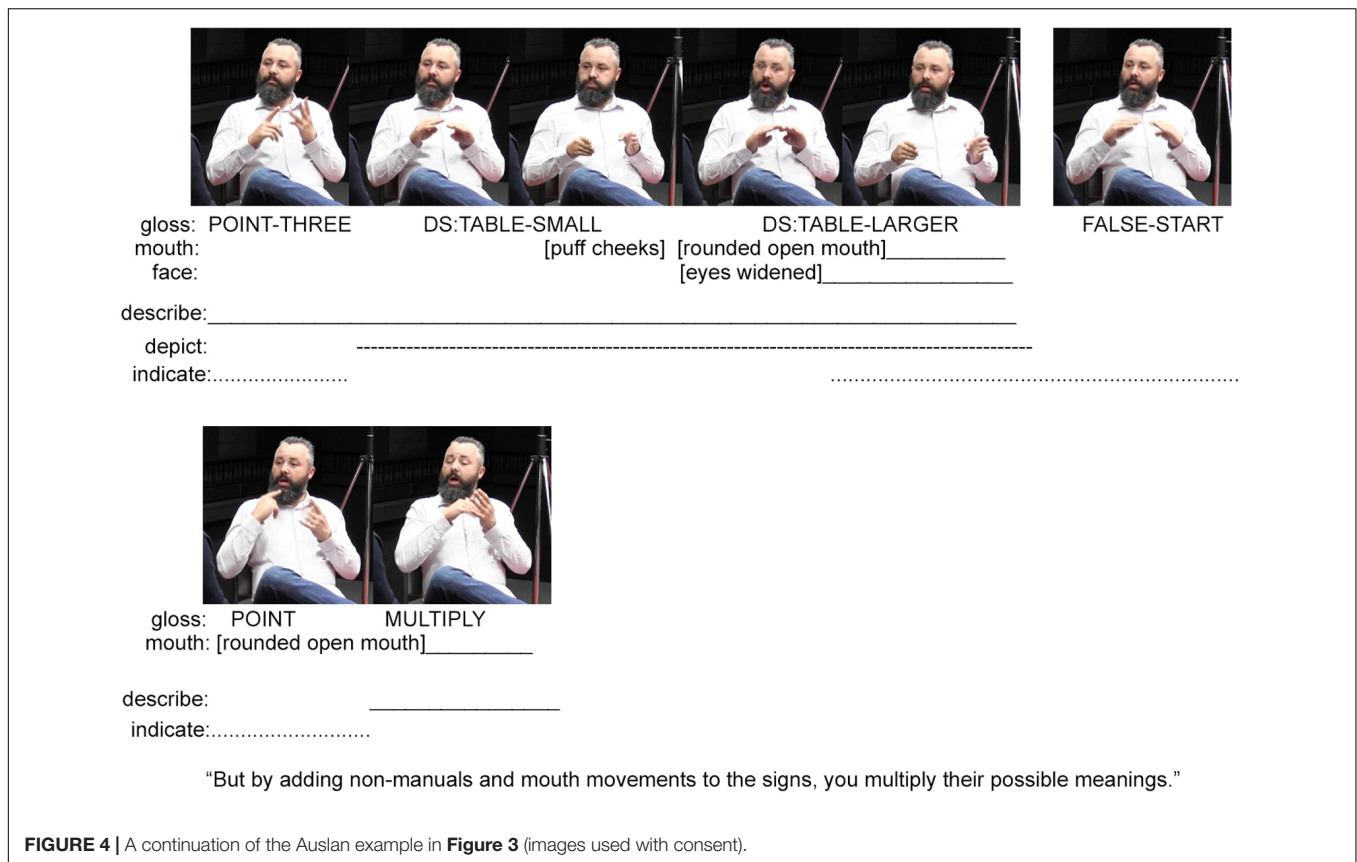
gloss: THREE SIGN THREE

mouth: 'three' 'sign'

describe: _____

“Like the signs for ‘small table’, ‘medium-sized table’, and ‘large table’ are three signs.”

FIGURE 3 | An example of an Auslan signer indicating, describing, and depicting across composite utterances (images used with consent, see Hodge et al., forthcoming for information about this dataset).



the alignment of speech with the enacted actions simultaneously makes these actions more precise (see the relevant discussion of this example in Kendon, 2004, p. 169).

The literature on spoken languages, signed languages, semiotics, gesture studies, and anthropology attests to a wide range of evidence for the ubiquity of different *P*-signs and composite utterances across varied communicative ecologies. For example, the use of co-speech pointing actions to symbolically index physical and abstract referents – and very often their simultaneous temporal and semantic alignment with speech – have been described for diverse language ecologies such as the Cuna people of Panama (Sherzer, 1972), the Yupno people of Papua New Guinea and speakers of American English (Cooperrider et al., 2014), Murrinhpatha in Northern Australia (Blythe et al., 2016), Kreol Seselwa in the Seychelles (Brück, 2016), and speakers of Nheengatú in Brazil (Floyd, 2016). Across these ecologies, pointing is both a plurifunctional and multimodal referential strategy (integrating bodily actions, posture orientations and eye gaze either with or without speech) that patterns along formal, semantic, and spatiotemporal lines.

Additional research into hearing speaker's use of co-speech gesture has shown that speakers' manual gestures offer either complementary or supplementary semantic information, or perform the same pragmatic function, as the spoken utterance (McNeill, 1992; Goldin-Meadow, 2003; Kendon, 2004; Calbris, 2011). Other manual gestures often co-occur with speech in various ways to achieve nuanced semantic understandings. Kita

and Özyürek's (2003) cross-linguistic comparison of speech and gesture ensembles produced during elicited narratives in Turkish, English, and Japanese found that speakers of all three languages consistently produce manual depictions of the same motion events. The exact manifestation of depicting actions varies between languages and appears to be shaped by grammatical structure (i.e., linguistic packaging), the lexical content of the speech utterance, and also spatial information in the elicited materials that was never expressed verbally in the speech acts. Loehr's (2012) analysis of the integration of intonation and manual gestures produced by English speakers indicates there is a strong temporal, structural, and pragmatic synchrony between speaker's speech and gestural production. For example, Loehr describes how one hearing English speaker uses manual gesture and a steep L + H* pitch accent to highlight a contrast between a present state being described and an earlier one (Loehr, 2012, pp. 84–85).

It has also long been observed that tokens of manual depictions or bodily enactments may replace constituent 'slots' in spoken composite utterances that are usually 'occupied' by conventionalized words (Slama-Cazacu, 1973; Kendon, 1988; McNeill, 2012). Slama-Cazacu (1973, p. 180) described this process as producing a "mixed syntax" within the interaction. Ladewig's (2014) research on manual gestures that replace speech within an utterance demonstrate how such gestures may function as verbs and nouns and are understood partly through the surrounding speech. She uses these observations as further

evidence that language is multimodal. Clark (2016) explains that depictions are a part of everyday utterances and that they may function as various types of constituents (e.g., a noun phrase, an object of a verb, a non-restricted relative clauses) or independently. The use of enactment in spoken language interactions has also been shown to co-occur and interact with the more conventional aspects of speech (Sidnell, 2006; Keevallik, 2018) – particularly when it is used for direct quotation (Bolden, 2004; Park, 2009; De Brabanter, 2010; Sams, 2010). Comparable patterns have also been described for signed language interactions (e.g., Metzger, 1995; Quinto-Pozos, 2007; Cormier et al., 2013b; Ferrara and Johnston, 2014).

Although not undertaken explicitly using a composite utterance approach, one investigation of clause structure in FinSL (Finnish Sign Language) found that deaf signers use variable constituent order and frequently omit overt argument expression from their utterances (Jantunen, 2008). Jantunen (2008, p. 112) also identified ample evidence of “important pantomimic aspects,” i.e., enactment, which could not be handled in existing frameworks for analyzing clause structure. Indeed, corpus-based analysis of the clause-like composite utterances in elicited retellings by deaf signers of Auslan has shown that tokens of enactment are frequently and tightly integrated into Auslan syntax at the clause level, e.g., a token of enactment may function as a core predicate constituent. Signers also use enactment to elaborate aspects of their narratives that are encoded lexically and may even rely solely on enactment to show and infer semantic relations between participants and events in a story, instead of explicitly encoding these relations via fully lexicalized manual signs and other conventionalized strategies of morphosyntax (Ferrara and Johnston, 2014; Hodge and Johnston, 2014). In some ways, these findings mirror findings on the integration of enactment and gesture in spoken language discourse mentioned above.

Investigations of BSL and Auslan have found that signers typically frame their enactments with lexical noun phrases and/or pointing actions, which function to index the referent subsequently enacted with the signer’s body (Cormier et al., 2013b; Ferrara and Johnston, 2014). Ferrara (2012) analyzed more than 5,000 composite utterances containing depicting signs produced by Auslan signers during elicited retellings and conversational activities. She found that these tokens of partly lexical signs often combined with other types of signs, but could also stand alone as full utterances. Another corpus-based analysis of approximately 1,000 clause-like composite utterances produced by Auslan signers during elicited retellings found that one in three tokens of core argument or predicate expression in single, stand-alone utterances was a partly-lexical pointing or depicting sign, or a token of enactment (Hodge and Johnston, 2014). More recently, Janzen (2017) has discussed topic-comment constructions and perspective-taking constructions (i.e., character viewpoints versus signer-as-narrator viewpoints) in American Sign Language (ASL) as composite utterances.

These studies illustrate how some patterns of argument structure and multimodal utterance composition constitute strategies of situated co-construction that emerge as the interactions unfold, and are therefore highly dependent on the

spatiotemporal context for recognition and interpretation. Given the essential role that indicating and depicting plays in signed interactions, these methods of signaling must be accounted for in signed language theory – as indeed they have been, albeit in various ways. We have seen that speakers also engage these methods of signaling. Thus, as signers and speakers both integrate multimodal indications and depictions into their utterances alongside descriptions in fairly conventional ways, a robust theory of language must be able to account for all three methods of signaling, even though token forms may vary in degree of conventionalization and how they are expressed across various language ecologies.

In the following sections, examples of composite utterances from deaf and hearing interactions are presented and discussed. First, two brief examples from interactions between deaf people are presented to illustrate how signers coordinate different types of *P*-signs within signed composite utterances. We then present an extended example that shows how deaf signers describe, indicate, and depict across longer stretches of interaction. In later sections, these examples are compared with examples from interactions between hearing speakers. Our aim is to demonstrate the coordinated signaling of description, depiction, and indication evidenced in both signed and spoken language interactions and achieve comparable analyses for both. We argue that Clark’s theory of language use is a strong starting point for uniting the communicative practices emerging within diverse ecologies under one theory of language. In this way, we extend Clark’s theory of language use to a theory of language.

COMPOSITE UTTERANCES EVIDENCED WITHIN DEAF/DEAF INTERACTIONS

A first example of a composite utterance evidenced in a deaf/deaf interaction is produced by a deaf Auslan signer re-telling *Frog, Where Are You?* (Mayer, 1969) to another deaf signer (**Figure 1**). During the story, a little boy searches for his missing pet frog. In retelling one moment of the story, the signer produces a composite utterance that both depicts and describes the boy as he picks up a boot and looks inside it. The signer begins with an enactment of the boy holding something over his head (i.e., a depiction), using eye gaze and facial orientation to index an as-yet un-named referent to a specific location in the signing space. This enactment is followed by a fingerspelled English word (‘boot’) and the lexical Auslan sign *BOOT* (i.e., a description of the object held by the boy). The signer completes his move with another enactment of the boy holding up the boot and looking into it (again, simultaneously depicting the event and indexing referents within the event). In this way, the signer coordinates different acts of description, indication, and depiction to create a composite utterance recounting a moment in the boy’s search for the frog. The initial enactment is elaborated retrospectively through the description of the referent ‘boot’ in both English and Auslan. The second iteration of the enactment enables the signer’s interactant to once again perceive what happened, but with clarified knowledge about the imagined object the boy (or rather, the signer as boy) was holding.

In this composite utterance, the descriptions, indications, and depictions are essential to understanding the meaning. Without the depictions, for example, all that would remain is a (bilingual) description of the referent ‘boot,’ which does little to move the story forward. In this example, we see that the availability of bodily enactment precludes the need to formulate a description through fully conventionalized lexis and grammar. We contend that such practices, based in the essentially face-to-face nature of interaction, have been able to fundamentally shape the signed languages of deaf communities (Johnston, 1996).

A second example further illustrates the nature of signed language communication by detailing a composite utterance produced as part of an informal conversation between three deaf Norwegians (**Figure 2**). The signer has almost finished detailing a personal experience about her childhood. She recounts how her father would have to physically come and find her when she was out playing, because she could not hear his calls. Her utterance begins with the signs POINT FATHER, thereby naming ‘father’ as the actor referent. The pointing action serves to index her own father, as opposed to someone else’s. The signer then elaborates her father’s actions by exploiting the gradient properties of the fully conventionalized sign RUN to express how her father would have to run (and find her). Her skillful manipulation of this lexical sign has the effect of profiling both descriptive and depictive elements of her expression. She ends this composite utterance by enacting her father as he ran to her, reached out and physically took hold of her, thus also indexing her young self as a referent through eye gaze and meaningful use of space. This depiction (which essentially functions as a verb) is framed by the phrase that both indicates and describes her father as the actor referent. As in the Auslan example, these descriptions, indications, and depictions are all integral to the intended meaning and must be interpreted holistically. If we were to focus only on the most conventionalized aspects of this utterance, i.e., the descriptive signs FATHER and RUN, then we would be left with only a partial understanding and analysis.

These two brief examples illustrate how deaf signers produce descriptions, indications, and depictions through manual and non-manual actions within composite utterances to express complex meanings. These methods of signaling cannot be easily isolated or divided from each other: they must be accounted for as integrated signals. The processes of describing, indicating, and depicting can be further clarified by examining an extended interaction between two deaf signers conversing in Auslan, i.e., a sequence of communicative moves comprising an interactional event (**Figures 3, 4**). Both signers are teachers of Auslan in Melbourne, engaging in a metalinguistic discussion about the strategies signers use to exploit and expand the comparably small lexicons of signed languages. This example consists of five composite utterances over 8 s. It was documented during the conversation task session for the Auslan and Australian English Corpus (Hodge et al., forthcoming).

The signer begins by producing three modified iterations of the fully conventionalized sign TABLE. By manipulating the depictive characteristics of the symbol TABLE, i.e., the resemblance in shape to a prototypical table, each iteration differentiates three tables of different sizes (**Figure 3**). Signers

exploit the iconic nature of signs in such ways as to manipulate meaning construction, and in doing so, they profile the dual function of many signs as descriptions and depictions (see also the sign RUN in the Norwegian Sign Language example in **Figure 2**; Johnston and Ferrara, 2012; Ferrara and Halvorsen, 2017). Comparable manipulations of iconic words have been observed in spoken languages (e.g., Dingemanse and Akita, 2016; Dingemanse, 2017a), which points to interesting similarities and differences between signed and spoken language ecologies.

Although these are just three versions of the lexical sign TABLE, the signer further explains that with different non-manual actions, one can multiply the meanings of these three signs. He does this by first describing his previous actions as ‘three signs’ via the fully conventional lexical signs (THREE SIGN THREE) and mouthings of the conventional English words ‘three’ and ‘sign’ (also illustrated in **Figure 3**). Using his right hand, he then points to the sign THREE, which was preserved on his left hand (**Figure 4**). This is possible because signers can hold signs over periods of time, creating possibilities of future interaction with those signs as physical entities. Although speakers are unable to hold spoken words over time while also continuing to speak, they can produce manual gestures that they interact with as physical entities.² The signer’s point to the sign THREE indexes the three signs depicting the differently-sized tables produced earlier. He then repeats these depictions while adding various mouth movements and non-manual actions to this reproduction (see the top row in **Figure 4**). The signer concludes by explaining that these non-manual actions “multiply the meanings of signs” (thus justifying why deaf signed languages do not require extensive manual signed lexicons). This explanation is expressed through a pointing sign that indicates his mouth (and thereby indexes the various movements undertaken during the preceding depictions) and a description (the lexical sign MULTIPLY), which explains the multiplying effect such non-manuals have on the meanings of signs. Again, this example demonstrates how methods for description, indication, and depiction are integrated within composite utterances. By focusing on one method of signaling only, we are unable to account for the full expression of the utterance – too much would be left out.

The three examples presented in this section show that deaf signers make strategic choices during the co-creation of composite utterances. Face-to-face interaction allows for the extensive use of all three methods of signaling, but particularly promotes the use of methods for indicating and depicting. The availability of space in deaf signed language interactions, we have seen, means that signers often rely heavily on indication and depiction for meaning construction. This has implications for the use of descriptions as well as the development of the inventory of conventionalized symbols which emerge within ecologies that are primarily (or in the case of deaf signed interactions, exclusively) face-to-face. Thus, theories of language which account only for conventionalized symbolic forms and the descriptions that signal them are incomplete, while also hindering an accurate understanding of how description works in combination with the other two methods of signaling (see also Liddell, 2003,

²Our thanks to a reviewer for pointing this out.

p. 362). Furthermore, the research reviewed in earlier sections has illustrated how hearing speakers also engage all three methods of signaling. One possible way to unite this knowledge into a global theory of language is to extend Clark's theory of (spoken) language use to that of language more generally, thus integrating findings from signed language linguistics, gesture research, and other disciplines into linguistic theory. More importantly, we can begin to understand how diverse humans communicate with each other without drawing haphazard and somewhat arbitrary lines around what is 'linguistic' and 'non-linguistic.' In the next section, we examine some examples of composite utterances evidenced in spoken language interactions to further demonstrate this position.

COMPOSITE UTTERANCES EVIDENCED WITHIN HEARING/HEARING INTERACTIONS

In this section, we turn our focus to examples of composite utterances produced during interactions between hearing speakers. By contrasting the composite utterances produced during deaf/deaf interactions with those produced during hearing/hearing interactions, we can begin to consider exactly how the communicative ecologies of signers and speakers may shape their coordination of methods for describing, indicating, and depicting within composite utterances. Firstly, an example from the literature briefly illustrates how hearing speakers create semantic and structural synchrony within their multimodal composite utterances:

[1] Ideophones and co-occurring manual gesture integrated with Siwu speech utterances (Dingemanse, 2014, p. 392):

gɔ ɔ-nyà ɔ-sè ɔ-ā'-bo,
when 3sg-see 3sg-hab 3sg-fut-reach
gɔ ɔ-nyà ɔdi àra,
when 3sg-see 3sg-take things,
"So when he got there, when he undressed,

gɔ ɔ-nyà kùgɔ ɔ-nya, ↑↑walayayayaya↑↑
when 3sg-see how 3sg-see, idph.walayayayaya
just when he's about to – walayayayaya!" ((gestures waves of water passing over skin))

oh, ɔ-tsùè pelepelepelepele
oh, 3sg:pst-burn idph.completely
"Oh, he was scalded all over."


In Example [1], the Siwu speaker depicts what happened to the king during an unfortunate bath by using conventional and non-conventional ideophones (*walayayayayaya* and *pelepelepelepele*) and manual gesture, while also describing what happened using fully conventionalized Siwu words and grammatical constructions. There are also examples of deictic morphemes (ɔ) that indicate the king as referent.

Similarly, Green and Wilkins (2014, p. 252) investigated the alternate signed language practices used by Arandic speaking communities of Central Australia and found that speakers

A: 'Cuz you know when I was booking a flight to go from
describe: _____
indicate:


A: ... Frankfurt to Barcelona
describe: _____

B: Mmm
describe: _____



A: And I could pick either **Lufthansa** or a **Spanish airline**
describe: _____
indicate:

B: Yeah
describe: _____

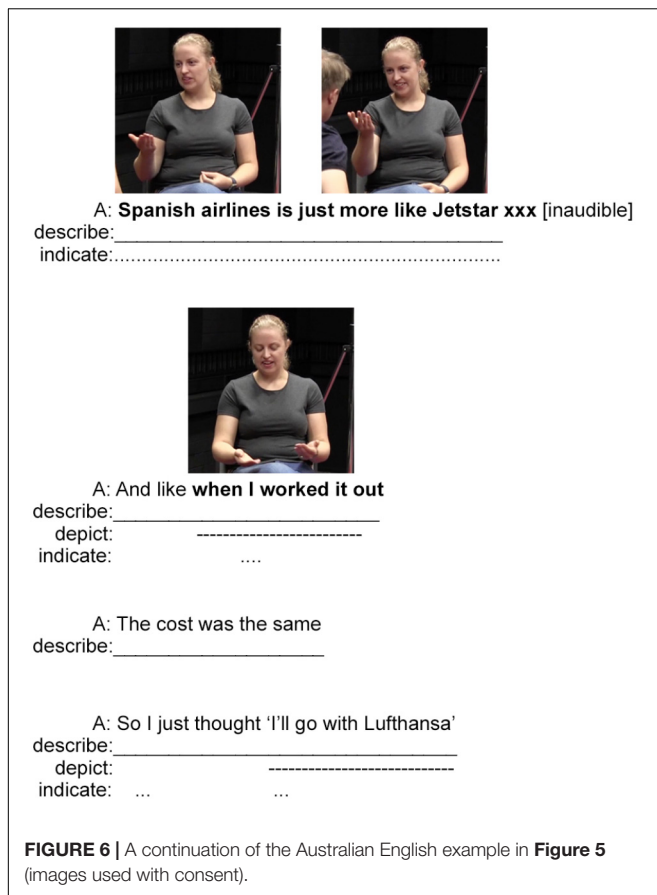


A: And I was like '**Lufthansa includes everything**'...
describe: _____
depict: _____
indicate: ...

FIGURE 5 | An example of an Australian English speaker describing, indicating, and depicting across composite utterances (images used with consent, see Hodge et al., forthcoming for information about this dataset).

habitually coordinate composite, multimodal packages with and without speech. These composite utterances involve different semiotic elements, including graphic depictions drawn in the sand, spoken words, and conventionalized signs produced with the hands, whereby each element serves to disambiguate the others. These patterns are akin to the ways in which Australian and Norwegian deaf signers use fully conventionalized signs and words to disambiguate their bodily enactments (see Composite Utterances Evidenced Within Deaf/Deaf Interactions). In each case, both signers and speakers make strategic, moment-by-moment choices about how to disambiguate the context of the interaction and prompt meaning construction, and then execute these choices by drawing from their available semiotic repertoire. A theory of language should be fully compatible with these choices by including both emerging and established communicative practices.

The next example involves composite utterances produced during an informal conversation between two hearing Australian English speakers. It was documented during the conversation task session for the Auslan and Australian English Corpus (Hodge et al., forthcoming). During this interaction, a hearing woman



is chatting to her brother about a previous experience booking a flight for travel in Europe. She explains how she compared two airlines and discovered that the low-cost airline was not so low-cost after all. She does this by coordinating her speech, hand, and body in acts of description, indication, and depiction. This example is presented in **Figures 5, 6** with relevant images of meaningful hand and body movements aligned with co-occurring periods of speech (represented in bold).

The example begins in **Figure 5** with two utterances that introduce the topic through a description using spoken English lexis and syntax: “Cuz you know when I was booking a flight to go from . . . Frankfurt to Barcelona.” The speaker then makes eye-contact with her brother (who was engaged with picking up a glass and taking a sip of water while she spoke) as he provides a confirmatory “Mmm.” She continues with a composite utterance that describes with speech the possible choice of two airlines. As she names the two airlines, she produces hand movements to indicate the two choices and locate the choices in space. These pointing actions also work to set the two choices up in opposition: she points her hands joined at the fingertips to the left of her leg to indicate Lufthansa, and then to the right to indicate a Spanish airline. This multimodal, composite utterance effectively presents the topic of conversation—namely a comparison of two airlines—through acts of description and indication.

In the next composite utterance, the speaker presents the first part of her comparison by combining speech, hand and head

movements, and facial expression to describe and depict her thought process (*And I was like “Lufthansa includes everything”*). The utterance begins with the English construction *And I was like*, which works to frame the subsequent depiction of (presumably) a thought process. The spoken part of this depiction is synchronized with the speaker raising her hands and shifting her gaze upward to demonstrate that the price from Lufthansa would be all-inclusive. Her hand movements in this utterance resemble what Kendon (2004) refers to as the Open Hand Supine gesture (OHS, in this case, a two-handed version), which has been analyzed as a gesture that relates to acts of receiving. Here, we can interpret this gesture as contributing to the meaning of the depiction that one receives everything included with a Lufthansa ticket, which may justify its higher initial price.

The speaker’s next composite utterance works to link the current interaction back to earlier comments her brother had made about the Australian low-cost airline Jetstar. She begins with a very brief manual indication to the Spanish airlines by producing another instance of an OHS gesture (this time on only the right hand) that she places on her right – notably, in the same space that indicated the Spanish airlines at the beginning of the example. Without directed movement, we may interpret this gesture as a Palm Presenting version of the OHS that presents the Spanish airlines as a focus. However, its function to indicate the Spanish airlines through meaningful location in space may mean this gesture is best analyzed as a Palm Addressed OHS gesture (see Kendon, 2004, Chapter 13). In any case, the gesture is accompanied by, and elaborated upon, with a description in spoken English, “*Spanish airlines*.” This phrase is followed by further description in spoken English that clarifies that the Spanish airline is similar to Jetstar. As the speaker utters this description, she once again produces an OHS gesture; this time a clearer example of the Palm Addressed type. She moves this gesture toward her brother, while also shaking the hand laterally, effectively acknowledging and indicating his earlier comments about Jetstar and their hidden costs.

The speaker then continues with two composite utterances that describe with speech the calculations she did to compare the costs between the airlines: “*And like when I worked it out, the cost was the same*.” While uttering these descriptions, the speaker also synchronizes a co-speech manual depiction comparing the two prices. This manual gesture can possibly be analyzed as depicting the ‘weighing of objects’ on a scale—the hands representing the surfaces of the two sides of the scale, which objects are placed upon [i.e., Müller’s (2014) representing gestures or Kendon’s (2004) modeling gestures]. An alternative analysis interprets the two hands as two calculations, again, representing gestures, that allow the speaker to visually inspect the choices. This example concludes with a framed depiction of the speaker’s final decision: *So I just thought “I’ll go with Lufthansa.”*

Overall, the acts of description, indication, and depiction coordinated within these composite utterances are very similar to the signaling acts produced during deaf/deaf interactions detailed in Section “Composite Utterances Evidenced Within Deaf/Deaf Interactions” and the hearing speaker in example [1] above. However, one difference between deaf/deaf and hearing/hearing interactions is immediately apparent: speakers recruit speech and

sound into their composite utterances in addition to visible bodily actions, whereas deaf signers typically only do this when they know the other person can hear. This fundamental difference reflects the respective lifeworlds and communicative ecologies of deaf and hearing people. The availability sound, or lack thereof, has important implications for analyzing and comparing signed and spoken interactions.

RE-ORIENTING LANGUAGE THEORY TO REFLECT MULTIMODAL LANGUAGE AS ACTION

In this paper, we have extended Clark's (1996) theory of language use to acknowledge that language and language use cannot be divided and to account for the diverse yet comparable communication practices which emerge during deaf/deaf and hearing/hearing interactions. As Dingemanse (2017b, p. 195) has commented, the tools we use to investigate language (i.e., our methods and theories) "enhance our powers of observation at one level of granularity (at the expense of others), and they bring certain phenomena in focus (defocusing others)." He suggests that sometimes these tools need to be recalibrated. In this paper, we have proposed re-calibrating traditional, structural theories of language with a more holistic theory that conceptualizes language as 'actioned' via three methods of signaling: describing, indicating and depicting. Evidence from the existing literature on signed and spoken languages demonstrates that these three methods of signaling are essential to understanding face-to-face communication. We have shown how both deaf signers and hearing speakers describe, indicate, and depict within composite utterances. In addition to signaling through description, both signers and speakers signal through indication and depiction within the spatiotemporal context of their unfolding interactions, although the exact manifestations of these patterns diverge according to the availability of

sound. These patterns attest to the pluralistic complexity of human communication and the varied semiotic repertoires which emerge within specific language ecologies. If we are to strive for robust and complex understandings of both signed and spoken language use, any language theory must acknowledge the broad range of intentionally expressive actions available to agents within specific spatiotemporal contexts, and the complex ecologies in which signers and speakers live. This can be achieved through direct comparison of the ways in which diverse humans produce and coordinate acts of description, indication, and depiction during their face-to-face interactions.

AUTHOR CONTRIBUTIONS

LF conceptualized the study, gathered the data, analyzed the data, and wrote 50% of the manuscript. GH detailed the theoretical argumentation, checked analysis of data, and wrote 50% of the manuscript.

ACKNOWLEDGMENTS

Documentation and early development of the Auslan and Australian English archive and corpus was supported by an Australian Research Council grant DP140102124 to Trevor Johnston, Adam Schembri, Kearsy Cormier, and Onno Crasborn. Additional corpus enrichment was supported by an ARC Centre of Excellence for the Dynamics of Language 2016 Language Documentation grant to GH. We are grateful to the Auslan signers and Australian English speakers who contributed to the archive, and the research assistants (Stephanie Linder and Sally Bowman) who worked to create it. The Norwegian Sign Language example used here comes from a pilot corpus project undertaken in 2015. We thank the Norwegian signers who participated in this project and graciously shared their language.

REFERENCES

- Agha, A. (2005). Voice, footing, enregisterment. *J. Linguist. Anthropol.* 15, 38–59. doi: 10.1525/jlin.2005.15.1.38
- Agha, A. (2007). *Language and Social Relations*. Cambridge: Cambridge University Press.
- Armstrong, D. F., Stokoe, W., and Wilcox, S. (1995). *Gesture and the Nature of Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620911
- Bavelas, J. B., Hutchinson, S., Kenwood, C., and Matheson, D. H. (1997). Using face-to-face dialogue as a standard for other communication systems. *Can. J. Commun.* 22, 5–24. doi: 10.22230/cjc.1997v22n1a973
- Bergman, B., and Dahl, Ö. (1994). "Ideophones in sign language? The place of reduplication in the tense-aspect system of Swedish Sign Language," in *Tense, Aspect and Action: Empirical and Theoretical Contributions to Language Typology*, eds C. Bache, H. Basbøll, and C.-E. Lindberg (Berlin: Walter de Gruyter), 397–422.
- Bernstein, B. (2003/1971). *Class, Codes and Control*. London: Routledge.
- Blythe, J., Mardigan, K., Carmelita, P., Mawurt, E., and Stoakes, H. (2016). Pointing out directions in Murrinhpatha. *Open Linguist.* 2, 132–159. doi: 10.1515/opli-2016-0007
- Bolden, G. (2004). The quote and beyond: defining boundaries of reported speech in conversational Russian. *J. Pragmat.* 36, 1071–1118. doi: 10.1016/j.pragma.2003.10.015
- Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.
- Bressem, J. (2013). "20th century: empirical research of body, language, and communication," in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and S. Tessendorf (Berlin: Mouton De Gruyter), 393–416.
- Brück, M. A. (2016). *Lalang, Zes ek Kiltir – Multimodal Reference Marking in Kreol Seselwa*. Ph.D. dissertation, Universität zu Köln, Köln. Available at: <http://kups.ub.uni-koeln.de/id/eprint/7964>
- Busch, B. (2015). Expanding the notion of the linguistic repertoire: on the concept of *spracherleben*—the lived experience of language. *Appl. Linguist.* 38, 340–358. doi: 10.1093/applin/amv030

- Bühler, K. (1990/1934). *Theory of Language: The Representational Function of Language*. Philadelphia, PA: John Benjamins.
- Calbris, G. (2011). *Elements of Meaning in Gesture*, trans. M. M. Copple. Amsterdam: John Benjamins. doi: 10.1075/g5.5
- Cienki, A., and Müller, C. (eds). (2008). *Metaphor and Gesture*. Amsterdam: John Benjamins. doi: 10.1075/g5.3
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620539
- Clark, H. H. (2003). "Pointing and placing," in *Where Language, Culture, and Cognition Meet*, ed. S. Kita (Mahwah, NJ: Lawrence Erlbaum Associates), 243–268.
- Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123, 324–347. doi: 10.1037/rev0000026
- Cogill-Koez, D. (2000). A model of signed language 'classifier predicates' as templated visual representation. *Sign Lang. Linguist.* 3, 209–236. doi: 10.1075/sll.3.2.04cog
- Cooperrider, K., Núñez, R., and Slotta, J. (2014). "The protean pointing gesture: variation in a building block of human communication," in *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, eds P. Bello, M. Guarini, M. McShane, and B. Scassellati (Austin, TX: Cognitive Science Society), 355–360.
- Cormier, K., Fenlon, J., and Schembri, A. (2015a). Indicating verbs in British Sign Language favour use of motivated space. *Open Linguist.* 1, 684–707. doi: 10.1515/opli-2015-0025
- Cormier, K., Quinto-Pozos, D., Sevcikova, Z., and Schembri, A. (2012). Lexicalisation and de-lexicalisation processes in sign languages: comparing depicting constructions and viewpoint gestures. *Lang. Commun.* 32, 329–348. doi: 10.1016/j.langcom.2012.09.004
- Cormier, K., Schembri, A., and Woll, B. (2013a). Pronouns and pointing in sign languages. *Lingua* 137, 230–247. doi: 10.1016/j.lingua.2013.09.010
- Cormier, K., Smith, S., and Sevcikova, Z. (2015b). Rethinking constructed action. *Sign Lang. Linguist.* 18, 167–204. doi: 10.1075/sll.18.2.01cor
- Cormier, K., Smith, S., and Zwets, M. (2013b). Framing constructed action in British Sign Language narratives. *J. Pragmat.* 55, 119–139. doi: 10.1016/j.pragma.2013.06.002
- De Brabanter, P. (2010). The semantics and pragmatics of hybrid quotations. *Lang. Linguist. Compass* 4, 107–120. doi: 10.1111/j.1749-818X.2009.00185.x
- DeMatteo, A. (1977). "Visual imagery and visual analogues in American Sign Language," in *On the Other Hand: New Perspectives on American Sign Language*, ed. L. A. Friedman (New York, NY: Academic Press), 109–136.
- Dingemanse, M. (2011). Ideophones and the aesthetics of everyday language in a west-African society. *Senses Soc.* 6, 77–85. doi: 10.2752/174589311X12893982233830
- Dingemanse, M. (2014). Making new ideophones in Siwu: creative depiction in conversation. *Pragmat. Soc.* 5, 384–405. doi: 10.1075/ps.5.3.04din
- Dingemanse, M. (2015). Ideophones and reduplication: depiction, description, and the interpretation of repeated talk in discourse. *Stud. Lang.* 39, 946–970. doi: 10.1075/sl.39.4.05din
- Dingemanse, M. (2017a). Expressiveness and system integration: on the typology of ideophones, with special reference to Siwu. *STUF Lang. Typol. Univ.* 70, 119–141. doi: 10.1515/stuf-2017-0018
- Dingemanse, M. (2017b). "On the margins of language: ideophones, interjections and dependencies in linguistic theory," in *Dependencies in Language*, ed. N. J. Enfield (Berlin: Language Science Press), 195–203.
- Dingemanse, M., and Akita, K. (2016). An inverse relation between expressiveness and grammatical integration: on the morphosyntactic typology of ideophones, with special reference to Japanese. *J. Linguist.* 53, 501–532. doi: 10.1017/S00222671600030X
- Dudis, P. (2011). "The body in scene depictions," in *Discourse in Signed Languages*, ed. C. Roy (Washington, DC: Gallaudet University Press), 3–45.
- Duranti, A., and Goodwin, C. (eds). (1992). *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge: Cambridge University Press.
- Emmorey, K. (ed.). (2003). *Perspectives on Classifier Constructions in Sign Languages*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511576737
- Enfield, N. J. (2013). *Relationship Thinking: Agency, Enchrony, and Human Sociality*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199338733.001.0001
- Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã: another look at the design features of human language. *Curr. Anthropol.* 46, 621–646. doi: 10.1086/431525
- Fenlon, J., Keane, J., Cooperrider, K. J., Brentari, D., and Goldin-Meadow, S. (2016). Comparing pronominal signs with pointing gestures. *Paper Presented at the International Society for Gesture Studies Conference, 18–22 July, Sorbonne-Nouvelle 3, Paris*.
- Fenlon, J., Schembri, A., and Kearsy, C. (2018). Modification of indicating verbs in British Sign Language: a corpus-based study. *Language* 94, 84–118. doi: 10.1353/lan.2018.0002
- Ferrara, L. (2012). *The Grammar of Depiction: Exploring Gesture and Language in Australian Sign Language (Auslan)*. Ph.D. dissertation, Macquarie University, Sydney, NSW.
- Ferrara, L., and Bø, V. (2015). *A Pilot Corpus of Norwegian Sign Language [Video Dataset]*. Trondheim: NTNU & HiOA.
- Ferrara, L., and Halvorsen, R. P. (2017). Depicting and describing meanings with iconic signs in Norwegian Sign Language. *Gesture* 16, 371–395. doi: 10.1075/gest.00001.fer
- Ferrara, L., and Johnston, T. (2014). Elaborating who's what: a study of constructed action and clause structure in Auslan (Australian Sign Language). *Aust. J. Linguist.* 34, 193–215. doi: 10.1080/07268602.2014.887405
- Floyd, S. I. (2016). Modally hybrid grammar? Celestial pointing for time-of-day reference in Nheengatú. *Language* 92, 31–64. doi: 10.1353/lan.2016.0013
- Fricke, E. (2014). "Deixis, gesture, and embodiment from a linguistic point of view," in *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and Communication Science 38.2)*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: Mouton De Gruyter), 1802–1823.
- Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help us Think*. Cambridge, MA: Harvard University Press.
- Goldin-Meadow, S., and Brentari, D. (2015). Gesture, sign and language: the coming of age of sign language and gesture studies. *Behav. Brain Sci.* 40:e46. doi: 10.1017/S0140525X15001247
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *J. Pragmat.* 32, 1489–1522. doi: 10.1016/S0378-2166(99)00096-X
- Green, E. M. (2015). "One language, or maybe two: direct communication, understanding, and informal interpreting in international deaf encounters," in *It's a Small World: International Deaf Spaces and Encounters*, eds M. Friedner and A. Kusters (Washington, DC: Gallaudet University Press), 70–82.
- Green, J. A. (2014). *Drawn from the Ground: Sound, Sign and Inscription in Central Australian sand Stories*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139237109
- Green, J. A., and Wilkins, D. P. (2014). With or without speech: Arandic sign language from Central Australia. *Aust. J. Linguist.* 34, 234–261. doi: 10.1080/07268602.2014.887407
- Harrelson, E. M. (2017). Deaf people with "no language": mobility and flexible accumulation in languaging practices of deaf people in Cambodia. *Appl. Linguist. Rev.* doi: 10.1515/applirev-2017-0081 [Epub ahead of print].
- Haugen, E. (1972). *The Ecology of Language: Essays*. Stanford, CA: Stanford University Press.
- Hodge, G. (forthcoming). "The ideology of communication practices embedded in an Australian deaf/hearing dance collaboration," in *Sign Language Ideologies in Practice*, eds A. Kusters, M. Green, E. M. Harrelson, and K. Snoddon (Berlin: Mouton de Gruyter).
- Hodge, G., and Johnston, T. (2014). Points, depictions, gestures and enactment: partly lexical and non-Lexical signs as core elements of single clause-like units in Auslan (Australian Sign Language). *Aust. J. Linguist.* 34, 262–291. doi: 10.1080/07268602.2014.887408
- Hodge, G., Sekine, K., Schembri, A., and Johnston, T. (forthcoming). Comparing signers and speakers: building a directly comparable corpus of Auslan and Australian English. *Corpora* 14.
- Jantunen, T. (2008). Fixed and free: order of the verbal predicate and its core arguments in declarative transitive clauses in Finnish Sign Language. *SKY J. Linguist.* 21, 83–123.

- Janzen, T. (2017). Composite utterances in a signed language: topic constructions and perspective-taking in ASL. *Cogn. Linguist.* 28, 511–538. doi: 10.1515/cog-2016-0121
- Johnston, T. (1996). “Function and medium in the forms of linguistic expression found in a sign language,” in *International Review of Sign Linguistics*, Vol. 1, eds W. H. Edmondson and R. B. Wilbur (Mahwah, NJ: Lawrence Erlbaum), 57–94.
- Johnston, T. (2008). “The Auslan archive and corpus,” in *The Endangered Languages Archive*, ed. D. Nathan (London: University of London).
- Johnston, T. (2012). Lexical frequency in sign languages. *J. Deaf Stud. Deaf Educ.* 17, 163–193. doi: 10.1093/deafed/enr036
- Johnston, T. (2013a). Formational and functional characteristics of pointing signs in a corpus of Auslan (Australian sign language): are the data sufficient to posit a grammatical class of ‘pronouns’ in Auslan? *Corpus Linguist. Linguist. Theory* 9, 109–159.
- Johnston, T. (2013b). Towards a comparative semiotics of pointing actions in signed and spoken languages. *Gesture* 13, 109–142. doi: 10.1075/gest.13.2.01joh
- Johnston, T. (ed.). (1989). *Auslan Dictionary: A Dictionary of the Sign Language of the Australian Deaf Community*. Sydney, NSW: Deafness Resources Australia.
- Johnston, T., and Ferrara, L. (2012). “Lexicalization in signed languages: when an idiom is not an idiom,” in *Proceedings the 3rd UK Cognitive Linguistics Conference*, Vol. 1, Hatfield, 229–248.
- Johnston, T., and Schembri, A. (2007). *Australian Sign Language: An Introduction to Sign Language Linguistics*. Cambridge: Cambridge University Press.
- Keevallik, L. (2018). What does embodied interaction tell us about grammar? *Res. Lang. Soc. Interact.* 51, 1–21. doi: 10.1080/08351813.2018.1413887
- Kelly, B. F. (2002). “The development of speech, gesture, and action as communicative strategies,” in *Proceedings of the 27th Annual Meeting of the Berkeley Linguistics Society: Language and Gesture* (Berkeley, CA: University of California).
- Kelly, B. F. (2006). “The development of constructions through gesture,” in *Constructions in Acquisition*, eds E. V. Clark and B. F. Kelly (Palo Alto, CA: CSLI), 11–25.
- Kendon, A. (1988). “How gestures can become like words,” in *Cross-Cultural Perspectives in Nonverbal Communication*, ed. F. Poyatos (Toronto, ON: C.J. Hogrefe Publishers), 131–141.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kendon, A. (2008). Some reflections on the relationship between ‘gesture’ and ‘sign’. *Gesture* 8, 348–366. doi: 10.1075/gest.8.3.05ken
- Kendon, A. (2014). Semiotic diversity in utterance production and the concept of ‘language’. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130293. doi: 10.1098/rstb.2013.0293
- Kita, S. (2003). “Pointing: a foundational building block of human communication,” in *Pointing: Where Language, Culture, and Cognition Meet*, ed. S. Kita (Mahwah, NJ: Lawrence Erlbaum Associates), 1–8.
- Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* 48, 16–32. doi: 10.1016/S0749-596X(02)00505-3
- Klima, E., and Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.
- Kockelman, P. (2005). The semiotic stance. *Semiotica* 157, 233–304. doi: 10.1515/semi.2005.2005.157.1-4.233
- Kral, I., and Ellis, E. M. (2008). “Children, language and literacy in the Ngaanyatjarra lands,” in *Children’s Language and Multilingualism*, eds J. Simpson and G. Wiggelsworth (London: Continuum), 154–172.
- Kravchenko, A. (2006). Cognitive linguistics, biology of cognition and biosemiotics: bridging the gaps. *Lang. Sci.* 28, 51–75. doi: 10.1016/j.langsci.2005.02.002
- Kusters, A. (2017). “Our hands must be connected”: visible gestures, tactile gestures and objects in interactions featuring a deafblind customer in Mumbai. *Soc. Semiot.* 27, 394–410. doi: 10.1080/10350330.2017.1334386
- Kusters, A., Spotti, M., Swanwick, R., and Tapio, E. (2017). Beyond languages, beyond modalities: transforming the study of semiotic repertoires. *Int. J. Multiling.* 14, 219–232. doi: 10.1080/14790718.2017.1321651
- Ladewig, S. H. (2014). “Creating multimodal utterances: the linear integration of gestures into speech,” in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and Communication Science 38.2)*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressen (Berlin: De Gruyter Mouton), 1662–1677.
- LaPolla, R. (2003). “Why languages differ: variation in the conventionalization of constraints on inference,” in *Language Variation: Papers on Variation and Change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, eds D. Bradley, R. LaPolla, B. Michailovsky, and G. Thurgood (Canberra, ACT: Australian National University), 113–144.
- LaPolla, R. (2005). “Typology and complexity,” in *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics*, eds J. W. Minett and W. S.-Y. Wang (Hong Kong: City University of Hong Kong Press), 465–493.
- Levinson, S., and Enfield, N. J. (2006). *Roots of Human Sociality: Culture, Cognition and Interaction*. Oxford: Berg.
- Liddell, S. K. (1995). “Real, surrogate, and token space: grammatical consequences in ASL,” in *Language, Gesture, and Space*, eds K. Emmorey and J. Reilly (Hillsdale, NJ: Lawrence Erlbaum Associates), 19–41.
- Liddell, S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511615054
- Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *J. Lab. Phonol.* 3, 71–89. doi: 10.1515/lp-2012-0006
- Maturana, H. R., and Varela, F. J. (1987). *The Tree of Knowledge: The Biological Roots of Human Understanding*. Boston, MA: Shambhala Publications.
- Mayer, M. (1969). *Frog, Where are You?* New York, NY: Dial Press.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.
- McNeill, D. (2012). *How Language Began: Gesture and Speech in Human Evolution*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139108669
- Metzger, M. (1995). “Constructed dialogue and constructed action in American Sign Language,” in *Sociolinguistics in Deaf Communities*, ed. C. Lucas (Washington, DC: Gallaudet University Press), 255–271.
- Mittelberg, I. (2008). “Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar,” in *Metaphor and Gesture*, eds C. Müller and A. Cienki (Amsterdam: John Benjamins), 115–154.
- Müller, C. (1998). “Iconicity and gesture,” in *Oralité et Gestualité: Communication Multimodale, Interaction*, eds C. Cavé, I. Guaitella, and S. Santi (Montréal/Paris: L’Harmattan), 321–328.
- Müller, C. (2014). “Gesture as deliberate expressive movement,” in *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon*, eds M. Seyfeddinipur and M. Gullberg (Philadelphia: John Benjamins), 127–151.
- Müller, C. (2016). “From mimesis to meaning: a systematics of gestural mimesis for concrete and abstract referential gestures,” in *Meaning, Mind and Communication: Explorations in Cognitive Semiotics*, eds J. Zlatev, G. Sonesson, and P. Konderak (Frankfurt: Peter Lang), 211–226.
- Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D., and Bressen, J. (eds). (2014). *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*. Berlin: De Gruyter Mouton.
- Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D., and Tessoroff, S. (eds). (2013). *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*. Berlin: De Gruyter Mouton.
- Neville, M., Haddington, P., Heinemann, T., and Rauniomaa, M. (eds). (2014). *Interacting with Objects: Language, Materiality, and Social Activity*. Amsterdam: John Benjamins.
- Okrent, A. (2002). “A modality-free notion of gesture and how it can help us with the morpheme vs. gesture questions in sign language linguistics (Or at least give us some criteria to work with),” in *Modality and Structure in Signed and Spoken Languages*, eds R. Meier, K. Cormier, and D. Quinto-Pozos (New York, NY: Cambridge University Press), 175–198.
- Park, Y. (2009). Interaction between grammar and multimodal resources: quoting different characters in Korean multiparty conversation. *Discourse Stud.* 11, 79–104. doi: 10.1177/1461445608098499
- Parmentier, R. J. (1994). *Signs in Society: Studies in Semiotic Anthropology*. Indianapolis, IN: Indiana University Press.
- Peirce, C. S. (1955). *Philosophical Writings of Peirce*. New York, NY: Dover Publications.

- Pennycook, A., and Otsuji, E. (2014). Metrolingual multitasking and spatial repertoires: 'Pizza mo two minutes coming'. *J. Sociolinguist.* 18, 161–184. doi: 10.1111/josl.12079
- Perniss, P., Özyürek, A., and Morgan, G. (2015). The influence of the visual modality on language structure and conventionalization: insights from sign language and gesture. *Top. Cogn. Sci.* 7, 2–11. doi: 10.1111/tops.12127
- Pickering, J. (1997). "Beyond cognitivism: mutualism and postmodern psychology," in *Brain, Mind and Physics*, eds P. Pykkänen, P. Plykkö, and A. Hautamäki (Amsterdam: IOS Press).
- Quinto-Pozos, D. (2007). Can constructed action be considered obligatory. *Lingua* 117, 1285–1314. doi: 10.1016/j.lingua.2005.12.003
- Sams, J. (2010). Quoting the unspoken: an analysis of quotations in spoken discourse. *J. Pragmat.* 42, 3147–3160. doi: 10.1016/j.pragma.2010.04.024
- Schembri, A., Jones, C., and Burnham, D. (2005). Comparing action gestures and classifier verbs of motion: evidence from Australian Sign Language, Taiwan Sign Language, and nonsigners' gestures without speech. *J. Deaf Stud. Deaf Educ.* 10, 272–290. doi: 10.1093/deafed/eni029
- Sherzer, J. (1972). Verbal and nonverbal deixis: the pointed lip gesture among the San Blas Cuna. *Lang. Soc.* 2, 117–131. doi: 10.1017/S0047404500000087
- Sherzer, J. (1991). The Brazilian thumbs-up gesture. *J. Linguist. Anthropol.* 1, 189–197. doi: 10.1525/jlin.1991.1.2.189
- Sidnell, J. (2006). Coordinating gesture, talk, and gaze in reenactments. *Res. Lang. Soc. Interact.* 39, 377–409. doi: 10.1207/s15327973rlsi3904_2
- Slama-Cazacu, T. (1973). *Introduction to Psycholinguistics*. Berlin: Mouton de Gruyter. doi: 10.1515/9783110800821
- Sperber, D., and Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.
- Streeck, J. (2009). *Gesturecraft: The Manu-facture of Meaning*. Amsterdam: John Benjamins. doi: 10.1075/gs.2
- Supalla, T. (1978). "Morphology of verbs of location and motion in American Sign Language," in *Proceedings of the 2nd National Symposium on Sign Language Research and Teaching*, eds C. Caccamise and D. Hicks (Coronado, CA: National Association of the Deaf), 27–46.
- Supalla, T. (2003). "Revisiting visual analogy in ASL classifier predicates," in *Perspectives on Classifier Constructions in Sign Languages*, ed. K. Emmorey (Mahwah, NJ: Lawrence Erlbaum Associates), 249–257.
- Sutton-Spence, R., and Woll, B. (1999). *The Linguistics of British Sign Language: An Introduction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139167048
- Tannen, D. (1989). *Talking Voices*. Cambridge: Cambridge University Press.
- Thompson, S. A., and Couper-Kuhlen, E. (2005). The clause as a locus of grammar and interaction. *Discourse Stud.* 7, 81–505. doi: 10.1177/1461445605054403
- Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130292. doi: 10.1098/rstb.2013.0292
- Wilcox, S. (2007). "Routes from gesture to language," in *Verbal and Signed Languages: Comparing Structures, Constructs and Methodologies*, eds E. Pizzuto, P. Pietrandrea, and R. Simone (Berlin: Mouton de Gruyter), 107–131.
- Wilkins, D. (2003). "Why pointing with the index finger is not a universal (in socio-cultural and semiotic terms)," in *Pointing: Where Language, Culture and Cognition Meet*, ed. S. Kita (Mahwah, NJ: Lawrence Erlbaum Associates), 171–215.
- Wilson, D., and Sperber, D. (1993). Linguistic form and relevance. *Lingua* 90, 1–25. doi: 10.1016/0024-3841(93)90058-5

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ferrara and Hodge. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Gesture and Sign: Cataclysmic Break or Dynamic Relations?

Cornelia Müller*

Faculty of Social and Cultural Studies, European University Viadrina, Frankfurt (Oder), Germany

The goal of the article is to offer a framework against which relations between gesture and sign can be systematically explored beyond the current literature. It does so by (a) reconstructing the history of the discussion in the field of gesture studies, focusing on three leading positions (Kendon, McNeill, and Goldin-Meadow); and (b) by formulating a position to illustrate how this can be achieved. The paper concludes by emphasizing the need for systematic cross-linguistic research on multimodal use of language in its signed and spoken forms.

Keywords: gesture and sign, McNeill's gesture-sign continua, multimodality of language use, singular gestures, recurrent gestures, silent gestures, emblems, conventionalization processes

INTRODUCTION

Throughout the relatively short history of gesture studies, the relation between gesture and sign continues to figure as a central topic. For sign language studies, the question has politically been a highly delicate one, and it remains a vital issue in contemporary sign language research. Fortunately, today, we are in a position to discuss the relation between gesture and sign against the solid background of sign language studies, leaving no doubts concerning the fundamentally linguistic nature of signed languages (Kendon, 2004, 2008, 2014; Steinbach et al., 2012; Goldin-Meadow and Brentari, 2017). Against this background, discussions of the relation between gesture and sign can be very openly reconsidered.

Recent contributions to this discussion are the paper by Susan Goldin-Meadow and Diane Brentari "Gesture, sign, and language: The coming of age of sign language and gesture studies," published in 2017, and Kendon's (2014) "Semiotic diversity in utterance production and the 'concept' of language." The two publications come to very different conclusions concerning the relation between gesture and sign. While Kendon's work on gesture and sign lays out a multitude of ways in which gestures and sign "are on common ground" (Kendon, 2004, chapter 15), Goldin-Meadow highlights differences between gesture and sign early on, postulating a 'cataclysmic break' between the two (Singleton et al., 1995). Informed by McNeill's theory of gesture, this idea involves a focus on spontaneously created gestures and on gestures that "predict learning" (Goldin-Meadow and Brentari, 2017, p. 1).

Kendon, on the other hand, used the term 'gesture' in a much broader sense. In his 2004 book, *Gesture: Visible Action as Utterance*, he suggests to use the term 'gesture' as "a label for actions that have the features of manifest deliberate expressiveness" (Kendon, 2004, p. 15; see Müller, 2014a for a minute appreciation of Kendon's notion of gestures as movements displaying deliberate expressiveness). In 2013 he suggests employing "utterance visible action" to refer to what is commonly referred to as 'gesture': "In this essay, I offer a survey of the main questions with which I have been engaged in regard to 'gesture,' or, as I prefer to call it, and as will be explained below, 'utterance visible action.'" (Kendon, 2013, p. 7). His work on 'utterance visible action' concentrates on hand movements, as indicated in an article the following year: "And although visible bodily

OPEN ACCESS

Edited by:

Marianne Gullberg,
Lund University, Sweden

Reviewed by:

Adam Kendon,
University College London,
United Kingdom
Glenn David McNeill,
University of Chicago, United States

*Correspondence:

Cornelia Müller
cmueller@europa-uni.de

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 15 January 2018

Accepted: 17 August 2018

Published: 10 September 2018

Citation:

Müller C (2018) Gesture and Sign:
Cataclysmic Break or Dynamic
Relations?. *Front. Psychol.* 9:1651.
doi: 10.3389/fpsyg.2018.01651

actions in the torso, head and face can and do play roles in what is said in an utterance, here I shall concentrate upon the way hand actions interact with what is spoken in the production of content.” (Kendon, 2014, p. 4). Kendon comes to the conclusion that the question of how gesture and sign relate must be shifted to “how visible bodily action is used in utterance construction” and “becomes as much a part of the study of speakers as, necessarily, it is already a part of the study of signers” (Kendon, 2014, p. 13).

Rather than starting with the current positions in that debate, this article offers a historical reconstruction of the discussion of the gesture-sign relation carried out in the field of gesture studies. Why it is useful to look back in close detail? We tend to assume, particularly in psychology and the cognitive sciences, that academic knowledge advances continuously, making publications quickly look ‘outdated’ and ‘overtaken’ by more recent ones. The underlying assumption is that more recent publications are the more knowledgeable and offer the most up to date state of the art of academic research. Sometimes, however, the most recent debates carry the burden of ‘older’ discussions – often implicitly. For the discussion concerning the relation between gesture and sign this is definitely the case. One goal of this paper is therefore to exemplify that a close reading of the history of a scholarly discussion may not only help evaluate current positions, but indeed may still offer valuable insights, ideas, concepts or analytical criteria to work with. Given the scope of this paper, the focus is on the discussion of the relation between gesture and sign as it was led in the field of gesture studies.

Note that what is presented here is a close reading of the writings of Kendon, McNeill, and Goldin-Meadow (and also her 2017 co-author Diane Brentari). Put differently, it is an analysis of their lines or argumentation as presented in the texts. It is not a reconstruction of their current opinions. It aims instead at presenting a history of the development of an academic discussion on the relation between gesture and sign from a point of view of linguistic gesture studies (Müller et al., 2013b). The paper thus presents the author’s view of the arguments. To substantiate this reconstruction, many quotations of the original formulations are included. The figures in the article are analyses of argumentations as reconstructed by the author.

The paper begins with a reconstruction of the relation between gesture and sign in three seminal strands of work: Kendon on gestures, visible actions as utterances, which include sign; McNeill and his reading of Kendon’s work, and his highly influential model of gesture-sign continua; and Goldin-Meadow’s idea of a cataclysmic break between gesture and sign. I show how Kendon’s work highlights commonalities, how McNeill underlines differences and discontinuities, and the grounds on which Goldin-Meadow comes to postulate a categorical divide between gesture and sign.

In the second part of the article, I draw on and develop Kendon’s work to counter Goldin-Meadow’s position. The theoretical framework against which this counter position

is formulated adopts a concept of language as inherently multimodal, usage-based, and dynamic (Müller, 2007a, 2008; Müller et al., 2013a) and assumes an understanding of gesture as deliberate expressive movement (Müller, 2014a). The term gesture covers the full spectrum of co-speech gestures: singular, recurrent, and emblematic (Müller, 2010, 2017). The three types of gestures differ with regard to forms and degrees of conventionalization and with regard to their typical linguistic and communicative functions. Singular gestures are created on the spot; although they are based on a culturally shared repertoire of techniques of gesture creation (e.g., gestural modes of representation Müller, 1998a,b, 2014b, 2016), the specific realizations in a given context are rather free and spontaneous. Recurrent gestures “merge conventional and idiosyncratic elements and occupy a place between spontaneously created singular and emblems as fully conventionalized gestural expressions on a continuum of increasing conventionalization,” and “involve emergent decompositions of gestural movements into smaller concomitant *gestalts*” (Müller, 2017, p. 276). Emblematic gestures are fully conventionalized gestural movements. Functionally, the three types of gestures differ in that singular gestures mostly serve ‘lexical’ functions, for instance as attributes (Fricke, 2013); recurrent gestures mostly serve pragmatic functions (Bressem and Müller, 2014a; Ladewig, 2014b), as do emblematic gestures (Teßendorf, 2013). However, while singular and recurrent gestures operate upon spoken language utterances (contributing semantically or meta-communicatively), emblems mostly realize full speech-acts, for example the ‘okay gesture’ (Müller, 2010, 2014c). These gestural speech-acts can entail vocalizations or sometimes be paralleled by a verbal speech-act (sometimes this is the case with insults). Often they are used to replace a spoken language utterance. The three kinds of gestures operate as prototype categories, that is, they are not separated by sharp boundaries, their relations are dynamic. Throughout this paper the terms ‘singular gestures,’ ‘recurrent gestures,’ and ‘emblematic gestures’ are used as meta-terms to keep track of the different ways in which the term ‘gesture’ is used in the various frameworks discussed. Against this theoretical background, dynamic relations between gesture and sign are discussed. Such relations concern (a) historical change from gesture to sign, and (b) synchronic comparison of spoken and signed languages. The former includes lexicalization processes; the latter involves functional integration of gestures within a signed or spoken utterance (multimodal language use), and contact situations between spoken and signed languages (e.g., recurrent gestures used in Sign Language, or signing entering spoken languages).

The paper thus offers a framework against which the relations between gesture and sign can be systematically explored further by reconstructing the history of the discussion in the field of gesture studies and by formulating a position to illustrate how this can be achieved. The paper concludes by emphasizing the need for systematic cross-linguistic research on the multimodal use of language in its signed and spoken forms.

GESTURE-SIGN CONTINUA AND GESTURE AS UTTERANCE VISIBLE ACTION: THE DEVELOPMENT OF THE DISCUSSION IN GESTURE STUDIES

At least as far back as the Enlightenment we find reflections on the relation between gesture and sign (Copple, 2013; Kendon, 1975, 2002, 2004, chapter 3; Lane, 1979; Müller, 1998b, p. 51–53). One of the major reasons why this interest continues to motivate contemporary discussions is that the question of how gestures and signs relate to one another promises to provide insights into the nature and the origins of language itself (Kendon, 2002, 2008, 2014; McNeill, 2012, 2013; Wilcox, 2013).

A seminal moment in contemporary gesture studies was the publication of McNeill's provocative paper "So you think gestures are non-verbal" (1985) with which he challenged the, at the time, dominant assumption that gestures were to be considered as *not* being related to language proper. Gestures were considered to be part of non-verbal communication, clearly and fundamentally different from language. Social psychologists Ekman and Friesen (1969) had presented a classification of non-verbal behavior, conceiving of hand gestures as illustrators to the stream of speech. Drawing on psycholinguistic evidence Feyereisen (1987), Butterworth and Hadar (1989) suggested a fundamental difference between gestures and speech (for an overview see Hadar, 2013). McNeill countered this position and engaged in a lively controversy with the then prevailing understanding of gesture as unrelated to language (McNeill, 1985, 1987, 1989). The importance of this discussion for gesture studies cannot be stressed enough. McNeill prepared the ground for a psychological and linguistic perspective on gesture, and showed that gesture is a highly valuable object of study for both psychologists and linguists. With the advent of Cognitive Science in the 1980s and 1990s, his model of gesture and speech as forming one integrated system opened the doors for linguists to study gesture. McNeill's (1992) monograph *Hand and mind. What gestures reveal about thought* paved the way for gesture studies to emerge as a field. In McNeill's psychological model, *gesture and speech are two sides of language*, each reflecting fundamentally different forms of thought (imagistic vs. propositional), but both indispensable because their categorical difference drives thinking as people are speaking. Kendon also adopted a critical stance toward the idea of gestures as forms of non-verbal communication. Already by Kendon (1972) had demonstrated the intimate link between gesture and speaking, and showed (in Kendon, 1980d) that gesture and speech are "two aspects of the process of utterance" (see also Kendon, 1983). Kendon's work thus historically anticipated McNeill's. This is reflected in the manifold references to Kendon's work in McNeill's early writings on gesture.

Highlighting Commonalities: Gestures and Signs as Utterance Visible Actions (Kendon)

Kendon underlined the tight integration of gestures with speech in the process of utterance formation. In Kendon (1980c), he

showed that gesticulation units are temporally aligned with 'speech units' and must be considered "an alternate manifestation of the process by which 'ideas' are encoded into patterns of behavior which can be apprehended by others as reportive of those ideas. It is as if the process of utterance has two channels of output into behavior: one by way of speech, the other by way of bodily movement." (Kendon, 1980d, p. 218). In contrast to McNeill, however, Kendon's interest in gesture early on included conventionalized gestures, so-called 'emblems' (Efron, 1941; Ekman and Friesen, 1969; Kendon, 1981, 1984), or 'quotable gestures' (Kendon, 1992) and, with his move to Australia, signed languages increasingly attracted his attention (Müller, 2007b). In the early 1980s, he published a series of papers on a kinesic system, a village sign language, employed by the Enga community in Papua New Guinea. Those papers offer a minute analysis of the formational properties, the semiotic functioning, and utterance construction of the Enga sign language (Kendon, 1980a,b,c). What began with an elaborate analysis of the primary sign language of the Enga in Papua New Guinea led to a broad study of alternate sign languages employed by Central Australian Aboriginal speech communities (Kendon, 1988b). In the same year as Kendon's monumental work on Australian Aboriginal sign languages was published, a small book chapter appeared, which later inspired McNeill's formulation of "Kendon's continuum" (Kendon, 1988a, 2004, p. 104–106; McNeill, 1992, chapter 2). Kendon put forward arguments – historical, functional, and material (i.e., concerning the medium of expression) – in support of his view that "no sharp dividing line can be drawn between gesticulation that encodes meaning in a holistic fashion and gestures which, like so-called 'emblems,' are not shaped on the spur of the moment but follow an established form within a communication community, or which like the signs in a sign language, can be shown to be structured systematically out of recombineable [*sic*] elements and which do indeed refer to meaning units of great generality, as do words." (Kendon, 1988a, p. 134) Kendon considers both gesticulation and sign as *one gestural medium of expression* (Kendon, 2004, pp. 104, 307).

Historical Connections Between Gesticulation and Signs: Development as Lexicalization Process

By taking into account the full spectrum of gestures – including conventional and non-conventionalized kinesic forms – a historical process of sign formation from ad hoc created visual actions comes into view that can be described as a lexicalization process. In this process, gestures change over time, becoming increasingly stable, and may even develop into kinesic words, signs within a signed language. In his 1988 paper Kendon discusses different aspects of this process in a section entitled "Lexicalization of Gesture." He begins by introducing emblems as "the functional equivalent of a complete speech-act," a sbeing "standardized in form" and that "come to have meanings of much greater abstractness and generality" (Kendon, 1988a, p. 136). Concluding that "these forms are much more like words than anything we have heretofore considered," he moves

on to describe what happens when gestures become fully lexicalized:

“Gestures can become fully lexicalized when, for one reason or another, speech cannot be used for prolonged periods, but when nevertheless, all of the functions of spoken interchange are required. In these circumstances, where a spoken language is not available to create a context for gestural use, and where propositions must be exchanged as well as acts of interactional regulation, gestural units must be established that can serve, as words do, to refer to units of meaning that can be recombined to create complex signs with specific meanings.” (Kendon, 1988a, p. 136).

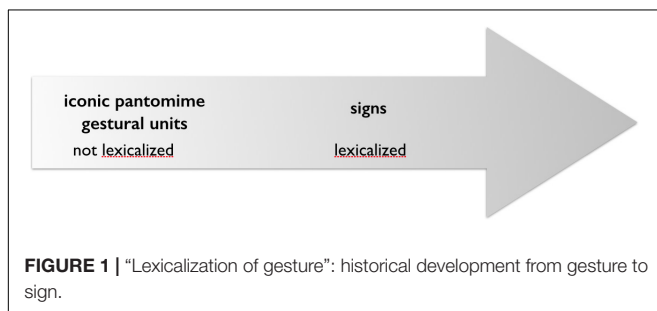
This historical-developmental perspective on the gesture-sign relation is illustrated in **Figure 1**, representing the analysis of Kendon’s text by the author of this article.

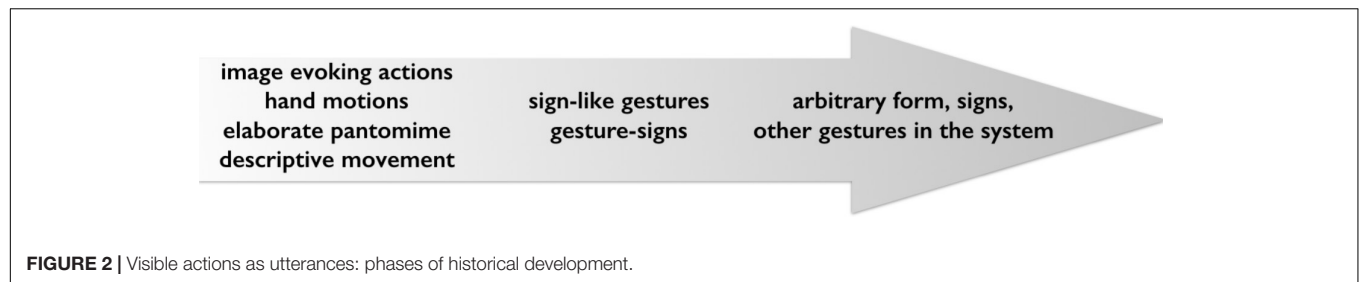
At that point, ‘gesture’ for Kendon includes the entire range of kinesic forms and functions, from gesticulation as spontaneously created form “that encodes meaning in a holistic fashion”, to emblems and, notably, it *includes* signs. Emblems differ from gesticulation in that they have acquired a fixed-form meaning relation. Kendon describes them as “following an established form” and as such they are comparable to words. In linguistic terms, these gestures are lexicalized. Signs are described as being “structured systematically out of recombineable [*sic*] elements and which do indeed refer to meaning units of great generality, as do words.” (Kendon, 1988a, p. 134). Signs within signed languages may result from processes of lexicalization that start from the ad hoc creation of kinesic forms, ‘gesticulation’. Kendon’s analysis of how gestures may become like words thus includes the development of gestures into ‘kinesic words.’ In his 2004 book, such a historical-developmental perspective is discussed under the heading of “Iconicity, sign formation and the emergence of ‘phonology.’” Here an example from Scroggs (1981) is reported that describes spontaneous creations of gestures of a deaf boy (not trained in sign language) which started as iconic pantomime and became increasingly reduced in form as the boy was telling his story. Kendon describes the process as beginning with “an elaborate pantomime of mounting the cycle, starting it, revving it up, using hand motions to indicate the twisting of the throttle on the handlebar” (Kendon, 2004, p. 308). Over the course of the story the pantomime becomes reduced and abstracted to a hand motion. In other

words, “the boy first created *representations* in gesture of the things he wished to refer to, and then he used *elements* from these representations as signs for these things.” (Kendon, 2004, p. 308; italics in the original). He then points out that it needs a speech-community for stable signs to develop and that the question of which elements are retained “in the transformation from elaborate depiction or enactment to a reduced sign-like gesture” depends upon their contrastiveness with “features of other gestures in the system” (Kendon, 2004, p. 308). Let me highlight two aspects of Kendon’s position as formulated here. First, Kendon speaks of ‘*sign-like gesture*,’ of ‘*gesture-signs*,’ and of ‘*other gestures in the system*’ using the term ‘*gesture*’ as a cover term for all kinesic forms of expression that are utterance dedicated visible actions used as utterances. Second, by describing the process of an emerging ‘gesture-sign,’ he spells out a *historical continuity between spontaneously created, singular forms of gesture (or gesticulation or descriptive movement) and simplified, standardized, arbitrary forms (signs) that function as words in a kinesic system*. Note that arbitrariness is considered to be an outcome of a historical process of change.

Kendon describes the phases of historical development as a path of transition that a spontaneously created gesture (gesticulation, singular gesture) may undergo on its way to an arbitrary sign (see **Figure 2**). He suggests that from “elaborate pantomime or descriptive movement sequence,” through simplification and “as a result of economy of action,” iconicity gets reduced (“is no longer apparent”) and “turns into an arbitrary form” under the ‘pressure’ to become a “distinctive form within a system of other forms” (Kendon, 2004, p. 308). Kendon not only points out that his view of the transitional process is grounded in his work on alternate and primary sign languages mentioned above, but also says, with reference to Klima and Bellugi (1979, chapters 1 and 3), Bellugi and Newkirk (1981), and Kyle and Woll (1985) that similar processes have been described in sign language studies *many times before*. In a nutshell, the argument Kendon unfolds is *an outline of the emergence of a kinesic language from spontaneous, singular forms of gestures, or from gesticulation*: “In this way, the visual representations and enactments for which the kinesic medium is so well adapted are transcended and a system of symbols that can operate in a quite abstract way is established.” (Kendon, 2004, p. 309).

Kendon’s ideas resonate with observations from (cognitive) linguistic research on historical changes of signs, here discussed under the label of *lexicalization and grammaticalization* processes. Wilcox (2005, 2007) describes routes from gestures to signed language with reference to American, Catalan, French, and Italian Sign Languages and with reference to historical documentations of gesture in the Mediterranean region. Several overviews of grammaticalization in sign languages have been offered (Wilcox et al., 2010; Pfau and Steinbach, 2011; Van Loon et al., 2014), and Janzen (2012) also discusses lexicalization. Kendon’s work can be considered an anticipation of this line of research and may also have been an incentive for it.





Functional Commonalities Between Gestures and Words in Spoken Languages

Already in that brief 1988a book chapter, Kendon brings in a second line of argumentation concerning the relation between gesture and sign: functional commonalities between gestures and words in spoken languages. From the point of view of communicative function, gestures can be used like words. This applies to all gestural forms, be they spontaneously created, holistic ones, or emblematic ones. Kendon already argues that gestures may be integrated in the vocal utterance, and then take over the function of a word. Many examples of semantic and pragmatic integration of a broad variety of hand-gestures in vocal utterances can be found throughout Kendon's work. In the 2004 book, chapter 8 offers a series of ways in which gestures are deployed in the utterance; chapter 9 is devoted to "gesture and speech in semantic interaction"; chapter 10 shows how referential meaning of gestures is established and how this interacts with what is being said. Chapter 11 shows different forms of pointing gestures and how they work in conjunction with speech. Chapters 12 and 13 then discuss semiotic motivation and contexts of use of gestures with pragmatic functions and how they form gesture families. These chapters include accounts of gestures such as 'precision grip' gestures (otherwise known as the ring gesture) as well as open hand gestures and reconstructions of their functions: marking topic-comment or questions for the precision grip in combination with the open hand supine (Kendon, 2004, p. 262) are two examples. For all kinds of gestures, close analyses of their integration into the verbal utterance are given. One example used again in his 2014 paper is a speaker gesturally showing the size of cheese crates as he says "and they used to come in crates about as long as that" and outlining their shape while saying "and they were shaped like a threepenny bit at the ends" (Kendon, 2004, p. 166). Slama-Cazacu (1976) had already described this phenomenon as mixed syntax. More recently it has been described as simultaneous construction (Vermeerbergen and Demey, 2007), as multimodal grammatical integration (Fricke, 2013), as multimodal utterance (Ladewig, 2014a; Ladewig, in press), as composite signal (Clark, 1996; Engle, 1998); or as composite utterance (Enfield, 2009, 2013; Clark, 2016 for speakers; Janzen, 2017 for signers). This is how Kendon (2014) describes this kind of gesture-speech interaction: "In his words, thus, he talks about the length of the crates, and he describes the sort of shape they had, whereas his hand actions are now seen as showing the length and the shape. It is as if he is using his hands to draw sketches of the objects he is talking about and, by means of these sketches, he adds a kind of description,

allowing, perhaps, the nature of the objects to be envisaged in a more precise way than the verbal description by itself might allow. The total meaning of what he is now saying is a product of an interaction between the meanings of his verbal phrases and the manually sketched illustrations that go with them. This is an example of what Enfield [34] has called a composite utterance." (Kendon, 2014, p. 5) In short, gestures, understood as visible actions, can become functionally equivalents of spoken language 'words,' they can form composite utterances.

Commonalities of Kinesic Medium: Gesture and Sign Share the Medium of Expression

As a third commonality between gesture and sign, Kendon points out that both forms of expression are produced in the same kinesic medium: "Speakers' uses of kinesic actions and signers' uses of kinesic actions are cut from the same cloth" (Kendon, 2004, p. 324, chapter 15). Given Kendon's intimate knowledge of primary and alternate sign languages and his work on conventionalized as well as non-conventionalized gestures, it is not surprising that material commonalities between gesture and sign come into view. In his 2004 book, an entire chapter is devoted to illustrating various ways in which 'gesture' and 'sign' can be understood as being 'on common ground.' Two issues are addressed: iconicity – involving sign formation and the emergence of kinesic phonology – and discourse construction. The discussion of iconicity and the emergence of kinesic phonology concerns the historical development of signs from spontaneously created gestures that we have dealt with above. Under the rubric of discourse construction, Kendon (2004, p. 310) discusses "features of the syntactic use of space and the use of 'classifiers' in sign language and describes examples of gesture use by speakers that seem very similar." Regarding the use of space, he suggests that speakers employ space in much the same manner as signers do. One example he gives compares the spatial inflection of signs as described by Liddell (2003), where signers set up so-called surrogate spaces to which they then point to later on in their discourse. Kendon gives examples where a speaker does just the same thing, first when setting up a gesture scene and later on pointing to the location set up before gesturally. Concerning sign language classifiers he suggests that they have much in common with what has been described of techniques of representation in gesture studies: "In American Sign Language there is a high degree of consistency in how the various hand shapes for the different classifiers are used, and how the movement patterns are carried out when they are employed. However, this seems to be but a regularization of techniques that

are widely used by speakers when using gesture for depictive purposes.” (Kendon, 2004, p. 318–319).

Utterance Visible Bodily Action: No Categorical Difference Between Gesture and Sign

Kendon offers three lines of argument in support of a view that sees no categorical difference between gesture and sign. He sees commonalities between gestures and signs with regard to historical, functional, and material aspects. In fact the commonalities between the two are considered so strong that he suggests giving up the term ‘gesture’ altogether and instead suggests replacing it by what he considers to be a more specific term: “utterance visible action” (Kendon, 2013, p. 7). He gives the following reasons for replacing the term gesture with “[...] utterance uses of visible bodily action”:

It is this that I shall call utterance visible action, and it corresponds to what is often referred to by the word “gesture.” However, because “gesture” is also sometimes used more widely to refer any kind of purposive action, for example the component actions of practical action sequences, or actions that may have symptomatic significance, such as self-touchings, patting the hair, fiddling with a wedding ring, rubbing the back of the head, and the like, because it is also used as a way of referring to the expressive significance of any sort of action (for example, saying that sending flowers to someone is a “gesture of affection”), and because, too, in some contexts the word “gesture” carries evaluative implications not always positive, it seems better to find a new and more specific term. (Kendon, 2013, p. 8).

In conclusion, Kendon’s position highlights *commonalities between different types of gestures and between gestures and signs*. In contrast to McNeill, he does not limit his account of the phenomenon to gesticulation (singular gestures), but includes conventionalized (recurrent) forms of co-speech gestures, emblematic gestures, as well as a thorough engagement with the analysis of sign languages.

Kendon (1988a) already suggested a bridge between gesture and sign against the backdrop of the historical development, functional and media specific commonalities:

I would like to suggest a different approach which, as I shall argue, can serve to link gesticulation with other kinds of gesturing, and which will also suggest that the gulf between presenting “content” in gesture and presenting it in “words” may not be as wide as it may now appear. At least I shall suggest a way in which a bridge may be built across that gulf. (Kendon, 1988a, p. 133).

The bridge Kendon offered a long time ago turned out to be not viable for McNeill and fellow psychologists, such as Singleton et al. (1995), Goldin-Meadow and Brentari (2017) or Emmorey (1999) (cf. also Kendon, 2000). Given their particular interest in gestures as windows onto thought, this is understandable. However, as we shall argue in the following section this comes at the cost of reducing the scope of gestural phenomena to those

kinds of gestures that are spontaneously created, that are global-synthetic, holistic in the McNeillian sense, that are capable of revealing the ‘imagistic’ thoughts of speakers (McNeill, 1992), and that are able to “predict learning” (Goldin-Meadow and Brentari, 2017, p. 1). *In short, it limits the study of gesture to one type, namely to singular gestures.*

The gesture studies community received Kendon’s (1988a) reflections on the relation between gesture and sign in terms of a gesture-sign continuum through McNeill’s discussion of it and through his (1992) formulation of ‘Kendon’s continuum’ as an interpretation of the positions Kendon had formulated (1988a). Kendon, however, never liked the term and asked McNeill to not use it, which McNeill followed in his 2000 revision of the original continuum (see also Kendon’s discussion of it under the heading “Kendon’s continuum revisited” in Kendon, 2004, chapter 6, p. 104–106). Quite surprisingly, McNeill introduced the term ‘continuum,’ but then used it to highlight *discontinuities* between gesture and sign. While at first sight this contradiction might not seem obvious, it is what McNeill’s reflections on the different ‘gesture-sign continua’ come to conclude. In fact, based on the discussion of a potential continuum between gesture and sign, McNeill *diagnoses a categorical difference between the two, a difference termed ‘cataclysmic break’* in a co-authored paper by Singleton, Goldin-Meadow and McNeill in 1995.

Highlighting Discontinuities: A Sharp Contrast Between Spontaneous Gestures and Socially Regulated Ones (McNeill)

It is puzzling. On the one hand, McNeill takes the radical counter position to psycholinguistic models on gestures by claiming that gestures are ‘verbal,’ meaning that they are an intrinsic part of language, rather than being *non-verbal*. On the other hand, he considers gestures as profoundly *different* from language. I propose that *this ‘difference’ is a consequence of a decision to restrict the concept of gesture to spontaneously used gestures.*

In McNeill’s work, the term ‘gesture’ refers only to singular gestures, gesticulation in Kendon’s terms. McNeill describes these gestural movements as being meaningful in a global-synthetic, holistic manner. McNeill (1992) clarifies that he uses the term “gesture” in this book specifically to refer to the leftmost, “gesticulation” end of the spectrum” (McNeill, 1992, p. 37). However, in ensuing discussions of the gesture-sign relation in the gesture studies community, the term ‘gesture,’ originally referring to singular gestures, came to be used as a cover term, *pars pro toto*, to refer to gestures in general. This led to a tacit backgrounding of recurrent and emblematic gestures that are nevertheless very widely used along with speech (Müller, 2017). While the palm-up-open-hand (PUOH) gesture is conceived of as a singular gesture, metaphorically presenting the topic of discourse (McNeill, 1992, p. 14–15; see Cienki and Müller, 2014, for a critical discussion of metaphoric gestures; but also Parrill’s, 2008 critique of the conventional status of the PUOH-gesture), other recurrent gestures are not systematically discussed. For McNeill, conventional gestural forms (recurrent and emblematic) were not in his focus of interest

since only spontaneously produced gestures (singular gestures) are psychologically interesting for him: they provide “an enriched view of the internal mental processes of speakers.” (McNeill, 1986, p. 108). They constitute a separate channel from speech and allow “a kind of triangulation onto the speaker’s mental representation” (McNeill, 1986, p. 108).

“A book about gestures and language.” This is how McNeill began his (1992) monograph. Crediting the discovery of the gesture-speech unity to Kendon’s observations on how gestures contribute to utterance construction, he had set out to develop a psychological theory of this relation. McNeill’s focus was always on singular gestures; as spontaneous creations of speakers they display individual ways of seeing the world. Singular gestures were viewed as images that are profoundly different from the conventional code of language, yet closely intertwined with speech:

The topic of this book was, specifically, gestures that exhibit images. With these kinds of gestures, people unwittingly display their inner thoughts and ways of understanding events in the world. These gestures are the person’s memories and thoughts rendered visible. Gestures are like thoughts themselves. They belong, not to the outside world, but to the inside one of memory, thought and mental images. Gesture images are complex, intricately interconnected, and not at all like photographs. Gestures open up a wholly new way of regarding thought processes, language, and the interactions of people.” (McNeill, 1992, p. 12).

It is important to go back to those very early formulations of McNeill’s theory of gesture and language, since they make it crystal-clear that he was interested in a specific kind of gestures, namely the individual, unique forms of gestures (i.e., the singular ones), because it is only these that allow insights into what he terms the imagistic side of language. This is the discovery he makes and he sets them apart from conventionalized gestures (recurrent and emblematic) that scholars from Antiquity to present times have dealt with: “None of these early investigators, however, considered the spontaneous gestures accompanying speech that are the chief focus of this book” (McNeill, 1992, p. 3). It is in the dialectic of singular gestures as ‘images’ and speech as a system of codified forms that McNeill sees two different forms of thought:

They [singular gestures] are closely linked to speech, yet present meaning in a form fundamentally different from that of speech. My own hypothesis is that speech and gesture [singular gestures] are elements of a single integrated process of utterance formation in which there is a synthesis of opposites modes of thought—global-synthetic and instantaneous imagery with linear-segmented temporally extended verbalization. Utterance and thought realized in them are both imagery and language (McNeill, 1992, p. 35).

This means, when formulating his hypothesis concerning speech and gesture as “elements of a single integrated process

of utterance formation” and characterizing this process as a “synthesis of opposites modes of thought—global-synthetic and instantaneous imagery with linear-segmented temporally extended verbalization,” singular gestures are being described as revealing the imagistic side of thought while speech reveals the linear-segmented form of thought. Put differently, what McNeill is interested in are the insights into ‘imagistic’ forms of thought that only the individual, spontaneously created gestures can offer.

This explains why *conventional (recurrent and emblematic) gestures are not in the scope of McNeill’s interest*. In his approach to gesture, conventional gestures switch sides, *they become like language and thus lose the unique capacity of opening up a window onto a speaker’s mind*. Conventional gestures are thus *qua definition* excluded from McNeill’s use of the term gesture. A continuum between the two thus cannot come into view, because these forms are excluded pre-hoc (as with emblems), or are not considered as being conventional (see above), which at least for the ‘ring gesture’ is undebatable even when used as a pragmatic co-speech gesture (Neumann, 2004; Müller, 2014c). The importance of the distinction between singular gestures and conventional recurrent and emblematic ones for McNeill is immense. He devotes the second chapter of his book to a substantiation of the fundamental difference between spontaneous gestures and codified signs:

The focus of this book is on spontaneous and idiosyncratic gestures (...) but it is useful to begin (...) with the more language-like gestures that constitute sign-languages. These are signs organized into true linguistic codes. *We benefit in this way from the sharp contrast that we can draw between the spontaneous and the socially regulated kinds of gesture.* (McNeill, 1992, p. 36; emphasis added).

The sharp contrast drawn by McNeill concerns singular gestures on the one hand, and recurrent and emblematic gestures on the other. In the formulation of this contrast, historical development and functional aspects are collapsed and put along one continuum, discussed broadly as gesture’s relation with speech (**Figure 3** adapted from McNeill, 1992, p. 3): “As we move from left to right: (1) the obligatory presence of speech declines, (2) the presence of language properties increases, and (3) idiosyncratic gestures are replaced by socially regulated signs.” (McNeill, 1992, p. 37) Note that here the term ‘gesture’ is used as a cover term to include spontaneous and conventional forms: gesticulation, language-like gestures, pantomimes and emblems.

What McNeill does is to put the functional integration of singular gestures into a verbal utterance (e.g., mixed syntax, or linear integration of ‘language-like’ gestures) on the same level as the historical development from gestures, to emblems, to signs. He thus blends the functional argument with the historical one. Moreover, in Kendon (1988a) gesticulation, language-like gestures and pantomimes are not described as alternatives. For Kendon gesticulation *includes* depictive as well as pantomimic gestures, and both can be used in a language-like function (**Figure 4**). But commonalities regarding gesture and sign as expressive medium are excluded from the continuum in McNeill.

Although McNeill later published a revised and expanded version of the continuum (McNeill, 2000), this blurring of historical and functional perspectives and the exclusion of commonalities concerning the kinesic medium of expression is maintained. Four aspects of the gesture-sign continuum are discussed separately:

(1) the relationship to speech, (2) the relationship to linguistic properties, (3) the relationship to conventions, and (4) the character of semiosis (McNeill, 2000, p. 1–7). **Figure 5** (adapted from McNeill, 2000) gives an overview of the changes along the continuum. Here again the term gesture is used in a



gesticulation language-like gestures pantomimes emblems signs

FIGURE 3 | McNeill's initial formulation of 'Kendon's continuum' (adapted from McNeill, 1992, p. 37).

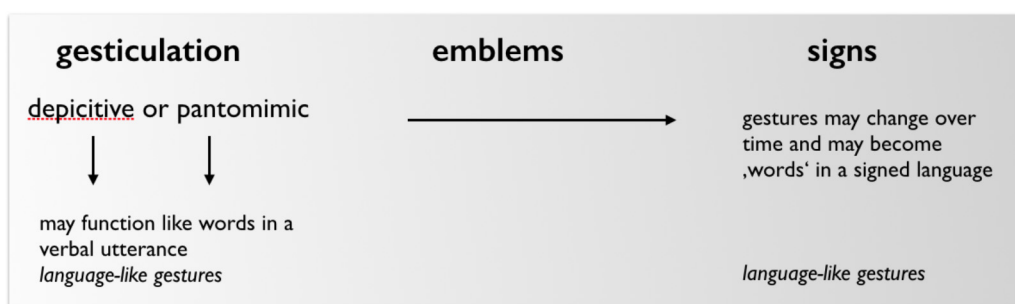
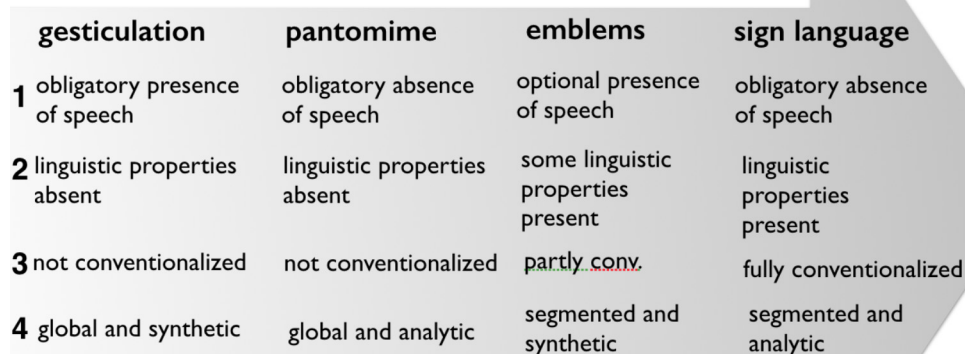


FIGURE 4 | Functional and historical aspects included in Kendon (1988a).



gesticulation	pantomime	emblems	sign language
1 obligatory presence of speech	obligatory absence of speech	optional presence of speech	obligatory absence of speech
2 linguistic properties absent	linguistic properties absent	some linguistic properties present	linguistic properties present
3 not conventionalized	not conventionalized	partly conv.	fully conventionalized
4 global and synthetic	global and analytic	segmented and synthetic	segmented and analytic

FIGURE 5 | McNeill's (2000) expanded version of the gesture-sign continuum (adapted from McNeill, 2000, p. 1–10).

broad sense to include non-conventional as well as conventional gestures (gesticulation, pantomime, emblems). **Figure 5** shows an overview of the four sub-continua.

McNeill suggests that, as one moves from gesticulation to sign, the obligatory presence of speech decreases (emblems and pantomime switch places here), linguistic properties (in terms of segmentation) increase, the character of semiosis changes from global-synthetic to segmented-analytic, and with conventionalization come emblems and signs. This description actually could be read as describing the historical processes of gesture change that both Kendon (1988a) and sign language studies describe as lexicalization (Janzen, 2012, see above and below). McNeill, however, establishes a clear-cut dividing line between gesture and sign as if processes of increasing conventionalization were impossible. Yet this is precisely what Kendon keeps pointing out. McNeill's continuum thus establishes a sharp dividing line between non-conventional and conventional forms *qua* an implied definition. *Instead of a gesture-sign continuum a categorical distinction between gesture and sign is established.*

But why are the continua so important for McNeill that he reconsiders them and even expands his exposition? The answer is that they are vital in defining the scope of phenomena covered by his psychological Growth-Point model. Only those forms of gesture that show an obligatory presence of speech, that have no linguistic properties, that are not conventionalized and whose meaning is constituted in global and synthetic manner (e.g., singular gestures) are able to reveal the imagistic side of thought. It is important to bear in mind that the concept of gestures as images is a rather idiosyncratic position of McNeill. Not only does it employ the term 'image' in a rather unelaborated manner, but it also backgrounds the fact that gestures are first and foremost *movements* of the hands often engaged in as-if actions and not images (Kappelhoff and Müller, 2011; Müller, 2014b, 2016, 2018). A concept of gesture as image disregards the practical engagements of the hands in mundane practices (cf., Streeck, 2009, 2013, 2017) and the way manual actions ground meaning of gestures (Müller, 1998a,b, 2004, 2010, 2014b, 2016, 2017; Kendon, 2004). For McNeill's model of thought processes, the *difference* between imagistic and propositional thought remains as fundamental as does the *difference* between spontaneous and conventional gestures, e.g., between singular and recurrent or emblematic gestures. It is the dialectic between imagistic and propositional forms of thought in the mental Growth-Point that, following McNeill, are said to drive thinking processes forward. When singular gestures become language-like, they change sides and also imprint thought with propositional structures which are characteristic of a conventionalized system of codified signs. That is, conventionalized gestures (recurrent and emblematic) are not in the scope of interest in McNeill's concept of gesture because only the individual spontaneous (singular) gestures of speakers reveal the hidden imagery of thought.

Such a limitation of the scope of phenomena under scrutiny is absolutely legitimate as long as it is dealt with explicitly, which McNeill very clearly does. It is very productive and even necessary for experimental studies. It is not helpful, however, in elucidating historical perspectives of gesture change,

commonalities between gestures and signs that concern their shared kinesic medium of expression, or the roles different types of gestures, including conventional gestures recurrently co-occurring with speech, play in the construction of multimodal utterances (Ladewig, 2014a,b,c; Müller, 2017; Ladewig, in press). Against this background, the postulation of a categorical divide or a 'cataclysmic break' between gesture and sign appears as deliberate exclusion of phenomena. This is perfectly legitimate to underline a specific aspect of gesture (as revealing spontaneous gestural forms of conceptualization, for example), or in an experimental design. It can, however, *not* be considered a response to the question of the gesture-sign relation in general, since it excludes recurrent and emblematic gestures, which are, nevertheless, extremely widespread aspects of multimodal utterance construction (Bressem and Müller, 2014a,b, 2017). The exclusion blurs potential 'continuities' that cannot come into view, since they fall outside the scope of the phenomena investigated. It also excludes reflections concerning the material commonalities between gesture and sign, relating to the medium of expression, both historically and when gestures are used by signers. Against this backdrop, the cataclysmic break recently restated by Goldin-Meadow and Brentari (2017) must be considered an '*artifact*' of definitions.

A 'Cataclysmic Break': 'Imagistic' Gesture and Categorical Sign (Goldin-Meadow and Brentari)

McNeill's position was formulated in the early nineties. It still informs the discussion on the relation between gesture and sign. Current discussions continue the above blurring of aspects of the gesture-sign relation that was an understandable consequence of McNeill's theory of gesture and language. In a recent paper, Goldin-Meadow and Brentari (2017) present a detailed overview of the state of the art concerning the relation between gesture, sign and language. It is a strengthening of the McNeillian position against a Kendonian view of that relationship. Goldin Meadow and Brentari's paper addresses the question: "How does sign language compare to gesture, on the one hand, and to spoken language on the other?" It tackles these questions strictly from a McNeillian point of view and "conclude that signers gesture just as speakers do. Both produce imagistic [singular] gestures along with more categorical signs or words." (Goldin-Meadow and Brentari, 2017, p. 1). The authors compare gesture-speech and gesture-sign systems as temporally co-existent systems, and apply the McNeillian concept of gesture as spontaneously created and 'encoding' meaning in an idiosyncratic, global-synthetic, holistic manner. In other words, they focus on singular gestures. Goldin-Meadow offers a psychological motivation for using the term gesture only for 'imagistic,' spontaneous forms of gesture: "we argue that making a distinction between sign (or speech) and gesture is essential to predict certain types of learning" (Goldin-Meadow and Brentari, 2017, p. 1). The authors conclude that "a full treatment of language needs to include both the more categorical (sign or speech) and the more imagistic (gestural) components regardless of modality and that, in order to make predictions about learning, we need to recognize (and figure out

how to make) a critical divide between the two” (Goldin-Meadow and Brentari, 2017, p. 2). It is crucial that Goldin-Meadow and Brentari make their definition of gesture explicit. What they do not make explicit, however, is that this excludes conventional co-speech gesturing once again, as in McNeill, *qua an implied definition*. As a consequence, what cannot come into view is a dynamic process of gesture change in which spontaneously created gestural forms may increasingly stabilize, and in which hybrid forms may emerge, such as is the case in recurrent gestures regularly employed by speakers across different discourse types (Müller, 2017, see also Kendon, 2004, p. 104). *The “critical divide” thus is a result, as in McNeill’s work, of the definitional limiting of gesture to singular gestures and the resulting disregard of conventional forms of co-speech gesture.* With regard to defining the scope of behaviors relevant for their work, Goldin-Meadow and Brentari apply the McNeillian framework and thus their position differs fundamentally from Kendon’s.

The different definitions or concepts of ‘gesture’ have important implications for the different concept of language the authors favor. While for Goldin-Meadow and Brentari, the simultaneity of spontaneous gesture with vocal and with signed languages shows the critical divide between what is ‘gesture’ and what is ‘language,’ for Kendon, the simultaneity of the full spectrum of ‘visible bodily actions’ with spoken and with signed languages indicates that the traditional concept of language is too narrow and should include the full range of visible bodily action as a close interrelation of different ‘semiotic systems.’ Kendon’s alternative to the concept of a sharp boundary between gesture and sign is the broadening of the concept of language to include different modalities and a flexible interrelation of different semiotic systems (Kendon, 2014, p. 3).

Goldin-Meadow and Brentari elaborate on the McNeillian position and highlight the endpoints of McNeill’s continuum to illustrate a discontinuity between gesture and sign. As a consequence, differences are maximized and the relation between singular gestures and signs is constructed as categorically distinct, as separated by a ‘cataclysmic break.’ This is how Singleton et al. (1995) formulated it in the title of a chapter: “The cataclysmic break between gesticulation and sign: Evidence against a unified continuum of gestural communication.” Here experimental evidence is offered to reject the idea of a continuity along the gesture-sign continuum that McNeill (1992) had attributed to Kendon’s (1988a) analysis. In a psychological experiment, speakers were placed in one of two conditions: describing previously seen events with and without speech. In the suppressed speech condition, the appearance of the gestures changed; they became more elaborate, more discrete. In the authors’ view, they became more language-like, more segmented, forming ordered strings. Goldin-Meadow and Brentari (2017, p. 9) summarize the results in the following way: “The gestures without speech immediately took on sign-like properties—they were discrete in form, with gestures forming segmented word-like units that were concatenated into strings characterized by consistent (non-English) order.” Notably, the authors attribute the change uniquely to the fact that spoken language was suppressed, and gestures had to carry the full communicative burden. The basic argument was to show that once an

individual had to communicate only manually, without making use of spoken language, the appearance of gestures changed *instantaneously*, and from one moment to another an individual speaker ‘invented’ signs. This might be why Goldin-Meadow (2015) characterized these gestures as ‘silent gestures’ and more recently as ‘spontaneous signs’ (Goldin-Meadow and Brentari, 2017). The implications drawn from this experiment are far-reaching and re-state a categorical divide between gesture and sign:

- (1) There is a qualitative difference between hand movements when they are produced along with speech (i.e., when they are gestures) and when they are required to carry the full burden of communication without speech (when they begin to take on linguistic properties and thus resemble signs); and (2) this change can take place instantly in a hearing individual. Taken together, the findings provide support for a categorical divide between these two forms of manual communication (i.e., between gesture and sign), and suggest that when gesture is silent, it crosses the divide (see also Kendon, 1988a). In this sense, silent gesture might be more appropriately called “spontaneous sign” (Goldin-Meadow and Brentari, 2017, p. 9).

If, however, the term gesture is reduced to singular gestures, then once again, the divide is caused by the definition. By deliberately excluding conventionalized co-speech gestures and by restricting the focus of analysis to a very specific experimental setting, gradual processes of change between spontaneous and conventional forms as they may happen in ordinary language use (Müller, 2017; see also Ladewig, 2010, 2011, 2014c) cannot come into view because of (a) *the definition of the term gesture*, and (b) *the restrictions of the experimental setting*. Consequently, conclusions drawn from this specific experimental condition are not viable for making claims beyond this specific experimental condition. Gesture change as an historical process can thus not come into view. This also holds for the various forms of functional integration of gestures within utterances as observed under naturalistic circumstances of language use. They are excluded, because they are not considered an object of inquiry.

Moreover, the interpretation of this experimental condition suggests that all it needs for language-like gestures to emerge is to suppress vocal language. However, no individual can produce a language. *What is needed for a language to appear is understanding, the reflexivity and intersubjectivity of meaning shared within a moment of discourse or across a community of speakers/signers.* Observing strings of ‘silent gestures’ under experimental conditions does not tell us whether they are understood by a conversational partner, or whether they function within a speech community. As a consequence, it does not tell us whether they *functionally* replace speech as a socially shared communicative system.

Goldin-Meadow and Brentari’s claim that all it needs for gesture to cross the ‘divide to language’ is to suppress spoken or signed language does not hold in light of observations concerning schematizations and generalizations of gestures and emergent sign described above. Kendon’s descriptions of processes of form

reduction and generalization of meaning happen very quickly in emergent signing and are extremely frequent in naturalistic contexts of multimodal language use (Kendon, 2013; see also Müller, 2017). Those processes obviously only can come into view under the condition (a) that the concept of gesture is not restricted to singular gestures, but includes singular, recurrent, as well as emblematic gestures, and (b) that gestures are studied across a broad range of different naturalistic discursive contexts.

Having introduced the idea of 'silent gestures' as indicators of a so-called cataclysmic break in the gesture-sign continuum, Goldin-Meadow and Brentari expand their view from the experimental setting to culturally shared repertoires of codified gestures arguing that those are the same kind of 'silent gestures' as the ones observed under the experimental condition described above. The common ground for those two very different forms of gesture *usage* is that they are said to be employed in the absence of speech. Put differently, the authors move directly from spontaneous co-speech gestures as produced under experimental conditions to codified sign systems. As a consequence, processes of gradual change cannot be uncovered, because precisely those kinds of gestures and those gestural usage contexts that could show such a gradual change are excluded.

However, the famous saw-mill gestures, monastic sign languages, or Aboriginal sign languages are all historical products of a communication community, they have evolved over time and have developed conventionalized repertoires of fixed form-meaning pairings, and a word-order (Kendon, 2004, chapters 14, 15; Kendon, 2013, p. 18). Kendon (2013) discusses these processes under the heading of "When utterance visible action is the main utterance vehicle." He points out that historical processes of sign formation have been widely discussed in sign language research that involve a historical and gradual transition from more complex kinesic enactments to more schematized ones, and this transition presupposes the social sharing of kinesic forms. Under naturalistic conditions of language use, it is through the back and forth between co-participants that schematization of forms and generalization of meaning happens (see also McNeill and Sowa, 2011):

To represent a meaning for someone else (and also, I think, to represent it for oneself), one resorts to a sort of re-creation. As if, by showing the other the thing that is meant, the other will come to grasp it in a way that overlaps with the way it is grasped by oneself. As these representations become socially shared, they rapidly undergo various processes of schematization. In consequence they are no longer understood only because they are depictions of something but also because they are forms which contrast with other forms in the system, acquiring the status of lexical items in a system. (Kendon, 2013, p. 18).

Rather than appearing instantaneously within one individual, codified kinesic languages are thus products of a historical process of language formation that critically depends on a community of users, be they engaged in a dyadic encounter or as members of larger communicative ensembles.

Although such a historical perspective on the gesture-sign relation clearly *contradicts* the discontinuity assumption of a 'cataclysmic break,' Goldin-Meadow and Brentari do mention processes of historical change: "Although the gesture forms initially are transparent depictions of their referents, over time they become less motivated, and as a result, less conventionalized, just as signs do in sign languages evolving in deaf communities (Frishberg, 1975; Burling, 1999)" (Goldin-Meadow and Brentari, 2017, p. 9). It is a logical consequence of their definition of gesture that, after conceding this historical process, the authors nevertheless come to the conclusion that "in all of these situations, the manual systems that develop look more like silent gestures than like the gestures that co-occur with speech." If the term gesture refers to singular gestures only (idiosyncratic gestures in McNeill's terminology and understanding) produced under experimental conditions, then (a) spontaneous processes of schematization and abstraction of singular gestures cannot come into view, because naturalistic conditions of use are not considered in which they happen very frequently, and (b) hybrid gesture forms that involve stabilized and non-stabilized formational aspects cannot come into view because recurrent and emblematic gestures are excluded qua definition (Müller, 2017).

Once again, the claim of a critical divide between gesture and sign is the result of a deliberate decision of (a) excluding conventional (co-speech and co-sign) gestures, and (b) experimental settings (which implies the exclusion of linguistic analysis of gesture-speech integration in its ordinary forms and contexts). Moreover, it implies a static and monadic concept of language as being either present or not, and as something that can appear 'instantaneously' within one individual.

Goldin-Meadow and Brentari also point out that 'silent gestures' in contrast to alternate sign languages, do not follow English word order. An explanation to this might be the fact that silent gestures are in fact, not like language at all. Because they lack the social sharing across a community of speakers and across the variable contexts of everyday life. The forms and repertoires of so-called 'silent gestures' never actually leave the experimental context, they are not taken up, changed, altered, adapted to other contexts of use, and they are never employed for complex communicative purposes. Thus, silent gestures do not have a chance to develop, simply because they are not used recurrently by a community of speakers under ordinary conditions of everyday life. Only if this happens, can we really see if English word order would be instantiated in gestures or not. For Goldin-Meadow and Brentari, these are the grounds on which they "argue that there are strong empirical reasons to distinguish between linguistic forms (both signed and spoken) and gestural forms," and "that doing so allows to us make predictions about learning that we would not otherwise be able to make." (Goldin-Meadow and Brentari, 2017, p. 2) Against our critical reading of the arguments, the empirical grounds presented by Goldin-Meadow and Brentari appear in fact rather weak. They rest upon (a) a restricted concept of gesture, (b) a highly specific experimental condition, and (c) a static and narrow concept of language. In fact, the narrow focus of their claims is asserted by the authors themselves, namely, by linking it to the possibility of making predictions about learning from singular gestures.

Clearly, adopting a narrow focus is legitimate for psychological research, and this is what they state in the above quotation, but three problems remain: (1) it does not tell us anything about how speakers and signers use recurrent and emblematic gestures; (2) it is not suited for proving a historical divide between singular, recurrent, and emblematic gestures and signs; (3) it does not tell us anything about functional commonalities between gestures and spoken or signed words.

Summing up, Goldin-Meadow and Brentari's position comes with a strong reduction of the scope of relevant behaviors included under the rubric of gesture, which clearly is crucial for psychological reasoning. For communicative, linguistic, anthropological, semiotic, and functional analyses of gestures this appears as a deliberate and artificial boundary which excludes *qua* definition hand movements in their full scope of phenomenological appearance in naturalistic settings. The validity of these findings for understanding relations between gestures and signs with respect to their communicative and linguistic functions must, therefore, be considered rather weak.

If the full spectrum of co-speech gesture is not considered, that is, *conventionalized* co-speech gestures are excluded, then gradual processes of change in the gestural medium of expression cannot come into view. What may happen if they are considered is the subject of the second section of this paper.

BEYOND THE CATACLYSMIC BREAK: DYNAMIC RELATIONS BETWEEN GESTURE AND SIGN

In this section, a plea is made for conceiving of relations between gesture and sign as dynamic. This shift involves a broad definition of the term gesture, and a consideration of gesture-sign relations from two different perspectives: the historical dynamics of gesture change, and a comparative view of two 'multimodal' languages in contact (for example, Deutsche Gebärdensprache, DGS, German Sign Language, and spoken German). The comparative perspective includes dynamic relations between gestures and signs within and across languages. It is informed by Kendon's multiple observations on the relation, as presented above, and it considers a discussion of gesture-sign continua as initiated by McNeill as vital for the discussion. The position sketched out here is thus informed by both lines of research in gesture studies. It does, however, not follow the assumption of a critical divide or a cataclysmic break between gesture and speech. Instead the relation between the two expressive modalities is considered as a dynamic one with regard to three different aspects: (a) historical development, (b) within, and (c) across spoken and signed languages. This position starts from a concept of language as inherently multimodal (Müller, 2007a, 2008). It is in line with Janzen (2017) who considers multimodality "a general characteristic of language, with composite utterances as instantiations of multimodality" (Janzen, 2017, p. 519). Furthermore, it is based on a linguistic perspective of multimodal language use (Müller, 2007a; Müller et al., 2013a; see also Ladewig, 2014a; Bressem and Müller, 2017; Ladewig, in press). It takes the analysis of multimodal language as

it is used across contexts as a basis for exploring manifold possible relations between gesture and sign (Müller, 2009; Bressem et al., 2018). This includes the analysis of gestures and signs across different naturalistic but also experimental contexts. It presupposes a close semiotic, interactional, and linguistic analysis of all the gestural forms we observe 'in the wild' (Müller, 2010, 2016, 2017; Bressem et al., 2013, see also Mittelberg, 2013, 2014; Mittelberg and Evola, 2014; Mittelberg and Waugh, 2014) and the multitude of ways in which they are integrated with speech or sign creating simultaneous structures (Vermeerbergen and Demey, 2007), composite utterances (Enfield, 2009, 2013; Janzen, 2017), gesture-speech ensembles (Kendon, 2004), or multimodal utterances (Ladewig, 2014a; Ladewig, in press). It also starts from a broader notion of the term gesture than the one suggested by McNeill and Goldin-Meadow.

Spelling Out the Concept of Gesture

Spelling out one's concept of gesture, even if an absolutely watertight definition remains unattainable, is crucial since it determines the scope of relevant behaviors that become relevant to empirical investigation and theoretical reflection. Moreover, it also explicates the theoretical framework within which a given assumption, research, proposal, and claims concerning gesture are formulated. As a consequence, the spectrum of phenomena covered by the claims is made explicit.

Although I agree with Goldin-Meadow and Brentari "that a full treatment of language needs to include both the more categorical (sign or speech) and the more imagistic (gestural) components regardless of modality (see also Kendon, 2014)" (Goldin-Meadow and Brentari, 2017, p. 2), I do not, agree with the assumption that gestural equals imagistic, nor that there is a clear-cut boundary between categorical and gestural.

From a usage-based and interactional point of view, gestures are meaningful body movements whose meaning is grounded in embodied experiences that are dynamic and intersubjective, and not at all like images (Müller, 2017; Müller and Kappelhoff, 2018). Put differently, I advocate an understanding of gestures as deliberate expressive movements (Kappelhoff and Müller, 2011; Müller, 2014a; see also Kendon, 2004, chapter 2). Semiotically, gestures are motivated by as-if actions, enactments of movement, or object representations (Müller, 2014b, 2016, 2017, 2018; see also Mittelberg, 2013, 2014). Gestures show degrees of conventionalization, understood as sedimentation of experiential frames (Müller, 2017). Degrees of conventionalization may range from none to partially to fully conventionalized. These different degrees are reflected in the terms "singular, recurrent, and emblematic gestures." Although the terms suggest categorical differences, these are not implied. Rather, we find different forms of hybridity between them (Müller, 2017).

An explication of the term 'gesture' helps to improve clarity in the discussion concerning the relation between gesture and sign. I favor using the term 'gesture' over the replacement 'utterance visible action' suggested by Kendon because, although this phrase was introduced to broaden the scope of behaviors under consideration, I suggest that, in fact, it narrows it down. Moreover, it implicitly establishes a specific theoretical focus. If 'utterance visible action' is applied *semiotically*, that is, if

it refers to the motivation of gestures, then this implies that gestures are *only* grounded in actions of the body. This excludes gestures that are enactments of movement and it excludes hybrid gestures, where some facets of a gestural movement, may be used to express aspects of meaning that are independent from the type of gesture. An example would be the deictic orientation of a horizontal ring gesture toward an addressee in contexts of expressing agreement and preciseness of an argument made by an interlocutor (a gestural expression of ratification and precision) (Müller, 2017). In that case the ring shape would be motivated by an as-if action of grasping while the movement toward the interlocutor is a deictic movement. Another case is the possibility to express aspectuality, understood as temporal contour of events, with a bounded (perfective) or an unbounded (imperfective) movement quality of a gesture (Müller, 2000, 2018). In a cross-linguistic study on aspectuality in Russian, French, and German significant correlations between perfective and imperfective past tense and bounded and unbounded gestural movement qualities were found for French speakers (Cienki and Iriskhanova, 2018). This perspective on the verbo-gestural expression of aspect goes along with a proposition made by linguists from various traditions (Behaghel, 1924; Holt, 1943; Croft, 2012), who proposed that verbs in the perfect(ive) tense characterize events as bounded in some way, as opposed to those in the imperfect(ive). Kinesically, boundedness was determined as pulse of effort, and unboundedness as more controlled movement, without a clear pulse of effort (Müller, 1998b, 2000, 2018; Boutet, 2010). We found that French speakers used significantly more “bounded” gestures, when they used the perfective tense (*Passé composé*). With the imperfective tense (*Imparfait*) the pattern was reversed. Speakers used more unbounded gestures (Cienki and Iriskhanova, 2018).

Furthermore, if ‘utterance visible action’ is understood as semiotic motivation in bodily actions only, then the concept would exclude gestures that are semiotically re-presentations of objects, when the hand becomes a body sculpture of a picture or a window or a piece of paper (Müller, 1998a,b, 2014b, 2016).

If, on the other hand, ‘utterance visible action’ refers to ‘action’ as a *theoretical concept*, then this implies a praxeological theory of communication (e.g., Streeck, 2013), which is an extremely important move in gesture theory interesting with far-reaching theoretical implication for gesture and speech as multimodal interaction, but it also implies a narrowing down of the theoretical focus more than the term gesture as currently employed in gesture studies. It is one of the strengths of the field of gesture studies that the term gesture allows for different theoretical frameworks to be applied and accordingly for different definitions and foci on gesture. As long as the respective concepts of gesture are spelled out explicitly, misunderstandings can be avoided and a critical discussion between the different positions fostered. To accurately gauge claims about gesture in Goldin-Meadow’s and McNeill’s work, it is important to know that the term gesture in their studies refers *only* to singular gestures (idiosyncratic gestures in the McNeillian sense). Conversely, to assess claims about gesture in Kendon, Streeck, or Müller’s work, it is equally important to know that here the term gesture involves a broader spectrum of bodily behaviors, including singular,

recurrent and emblematic gestures. Kendon’s recent plea for the notion of ‘utterance visible action’ obviously includes all of those and even “actions performed in the course of creating utterances in sign language” (Kendon, 2013, p. 8).

The following sections will illustrate how such a broad concept of gesture reveals dynamic relations between gesture and sign that a narrow one excludes qua definition.¹ The discussion and the claims made concerning the relation between gesture and sign are structured around two perspectives: a historical and a comparative one (within and across spoken and signed languages). Adopting such a broad concept reveals dynamic relations to be a fundamental characteristic of gesture.

Gesture Change: Historical Dynamics From Gesture to Sign

When taking into account the full spectrum of gestural expression, it becomes clear that non-stabilized, somewhat stabilized, and fully conventional gestural forms may be employed by language users. These forms are not sharply separated from one another as discrete categories. Rather, they can be thought of as arranged on a continuum from individually improvised forms to forms that are fully conventionalized. This is in line with Kendon’s (2014, p. 6) position: “These (and other) representational practices [...] are widely shared and are subject to varying degrees of social conventionalization. Some forelimb utterance actions may become so standardized that they acquire meanings that may be glossed with stable verbal expressions (often known as ‘emblems’ [39]), and, as such, are sometimes used as substitutes for spoken words in some contexts. In this case, we have something comparable to a lexical sign in a sign language” (Kendon, 2014, p. 6). However, in addition to Kendon’s sketch, we include recurrent gestures as an intermediate and hybrid form of gesture that is placed between singular and emblematic gestures with regard to conventionalization (Ladewig, 2010, 2011, 2014b,c; Müller, 2010, 2017). This developmental position between gesture and sign critically rests on their material commonality as a medium of expression. **Figure 6** systematizes a potential historical dynamics based on the degree of conventionalization and compositionality as an emergent feature. Note that historical development from gesture to sign may start with any of those three types of gestures.

Figure 6 is inspired by McNeill’s continuum (3) (McNeill, 2000, p. 4) and reflects an understanding of conventionalization as a successive, dynamic process of constant change (see also Gullberg’s, 1998 discussion of the continuum) and agrees with Gullberg’s refined systematics of the continuum’s left side, where she points out that the spontaneous forms of gesture (gesticulation) in fact entail a range of different varieties and includes, for instance, depictive as well as pantomimic forms (Gullberg, 1998, chapter 3). Singular gestures are considered to be gestures that are not conventionalized, that show a variable relation of form and meaning, and that are not compositional.

¹Note that although the author favors the use of the term ‘gesture’ over ‘visible action as utterance’, the position advocated stands in the tradition of Kendon’s work on the relation between gesture and sign, and was initiated and inspired by his work early on.

Figure 6 thus illustrates conventionalization as a gradual process (McNeill describes emblems as partly conventionalized, see also Gullberg, 1998, chapter 3), and introduces recurrent gestures: “By merging conventional with idiosyncratic or other conventional elements, recurrent gestures occupy a place between spontaneously created (singular) gestures and emblems as fully conventionalized gestural expressions on a continuum of increasing conventionalization” (Müller, 2017, p. 278). Examples of recurrent gestures are gestures that build families in the Kendonian sense, and that come with a stable form-meaning pairing (Bressem and Müller, 2014a,b, 2017; Ladewig, 2014b,c). A consequence of such conventionalization processes is that they affect gestural forms and functions gradually, and involve *hybridization* of spontaneous and more stabilized gestural forms and functions (Müller, 2017). From such a perspective, compositionality is a consequence of a process of decomposing holistic form-meaning units into stabilized formational cores with a shared semantic theme (to employ Kendon’s terms here, see also Kendon, 2004, p. 104). Formational features that are not involved can be used to express local meanings spontaneously (position in gesture space is often used in this way), or they can include other stabilized formational features. Recurrent gestures thus show emergent forms of compositionality. In emblematic gestures all formational features tend to be stabilized, and in that sense they are not compositional. Signs, however, are conventionalized and compositional, as in the case of spatial verbs described above. The compositionality of signs might be a consequence of accommodation and assimilation within a linguistic system, but this issue needs further exploration, at least as far as comparative gesture-sign language studies point of view is concerned. Recurrent gestures differ from emblems not only regarding their hybridity, but also regarding their functions (Müller, 2010, 2017; Ladewig, 2014b). Recurrent gestures function meta-communicatively and are thus inextricably connected with speech, emblematic gestures

are fully conventionalized and typically function as complete speech-acts; although they often include vocal elements, they are more independent from the co-presence of speech than recurrent gestures (Teßendorf, 2013).

When considering the full range of gestural phenomena, which, as we have seen, was Kendon’s position early on when he argued that gestures may lexicalize, it is possible to see that gestures are affected by processes of conventionalization, which – as in spoken language – are gradient and not at all sudden. Those processes go along with tacit agreements of a community of language users and the changes involved concern gestural forms *and* functions that emerge *from*, and change *with*, language use. In sign language research, such processes of change have been described in terms of lexicalization and grammaticalization, that is, as historical development from gesture to sign (Janzen, 2012, 2017):

Grammaticalization is the diachronic process by which lexical items develop into grammatical items in a language, or where items that are less grammatical in nature increase in their grammatical function (Heine et al., 1991; Bybee et al., 1994; Bybee, 2003; Hopper and Traugott, 2003; Brinton and Traugott, 2005; others). Grammaticalization in signed languages has been shown to develop by the same robust principles as for spoken languages with the exception that, whereas, for spoken language, historic sources for grammatical elements can only be shown to be earlier words, for signed language, grammatical elements can sometimes be traced back to gestural origins (Heine and Kuteva, 2007). Among such studies on ASL, Janzen (1998, 1999) has outlined the grammaticalization of topic marking as developing from a generalized questioning gesture, through regularized yes/no question marking, to topic marking. Janzen (1995) shows the development of the ASL lexical verb FINISH into both perfective and

singular gestures	recurrent gestures	emblems	signs
not conventionalized	partly conventionalized	fully conventionalized	fully conventionalized
no compositionality	emergent compositionality	no compositionality	compositionality
variable relation of form and meaning	stabilization of form-meaning core	fixed form-meaning unit	fixed form-meaning unit

FIGURE 6 | Gesture change: historical development from gesture to sign in terms of degree of conventionalization and compositionality as an emergent hybridization.

completive markers. Wilcox and Wilcox (1995), Shaffer (2000, 2002, 2004), Janzen and Shaffer (2002), and Shaffer and Janzen (2016) have outlined a number of ASL modals that have gestural sources for their development, and the evolution of discourse markers has been undertaken by Wilcox (1998)” (Janzen, 2017, p. 516–517).

Summing up, from a historical point of view, we observe gesture changes that are comparable to language change: *a historical dynamics of gesture and sign*. Gestural forms may stabilize (through repeated usages) and in some cases, undergo processes of lexicalization and grammaticalization and transform into signs within a signed language.

Dynamic Relations Across Languages: Comparing Co-speech Gesture to Co-sign Gesture

Karen Emmorey’s provocative paper “Do signers gesture?” diagnosed a discontinuity assumption concerning the relation between co-speech gesture and co-sign gesture (Emmorey, 1999). In a recent discussion of this question, Janzen points out that although Emmorey discusses commonalities between co-speech and co-sign gestures, she concludes “that essentially a signer’s gestures are not like co-speech gestures” (Janzen, 2017, p. 514). Kendon (2004, p. 324) also underlines that Emmorey’s paper insinuates a sharp distinction between gesture and sign, while at the same time providing examples of “how signers may insert ‘gestures’ into their discourse” (Kendon, 2004, p. 324).

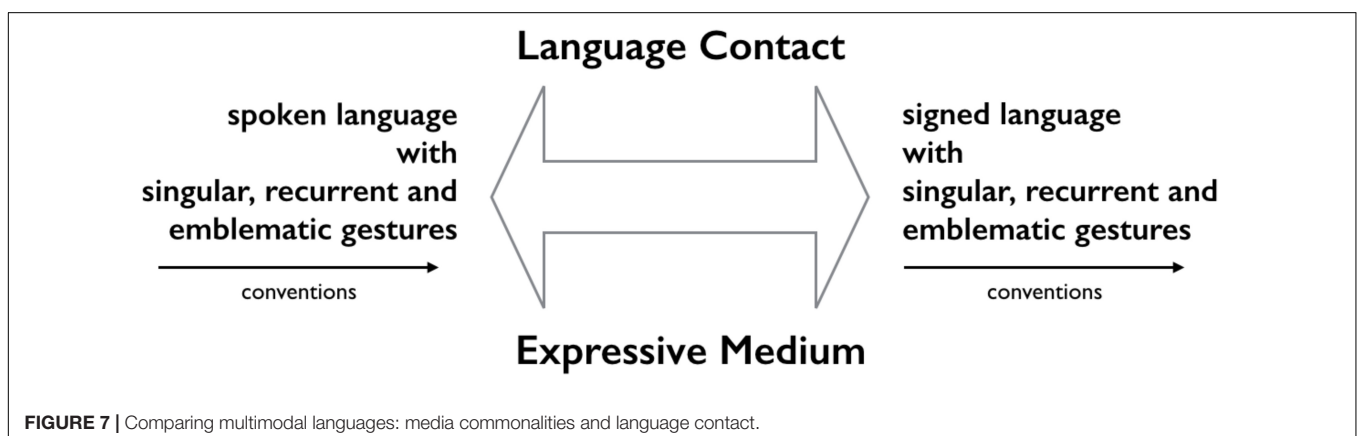
Liddell has argued that the ASL use of space, depicting verbs, pointing, and listing buoys is gestural (see Janzen, 2017, p. 515–518, and Kendon, 2004, p. 310–311 for discussions of this work). Janzen points out that while for Liddell “it is clear that he considers gestural material to exist pervasively in modern ASL” (Janzen, 2017, p. 516), nevertheless a sharp boundary between gesture and sign is established (Janzen, 2017, p. 518). Commonalities between co-speech and co-sign gestures have furthermore been discussed in the context of constructed or depicted action (Dudis, 2011; Janzen, 2017, p. 527; Hübl and Steinbach, 2018), classifier constructions (Supalla, 1982, 1986), or nominal proforms (Schembri, 2003; see Janzen, 2017, p. 525–526 and Kendon, 2004, p. 316–324 for detailed expositions).

Vermeerbergen and Demey (2007) suggest a simultaneity of gestures with spoken and signed language, and Goldin-Meadow and Brentari (2017, p. 1) “come to the conclusion that signers gesture just as speakers do.” Janzen proposes a usage-based, discourse-led approach to language as multimodal, as resting upon composite utterances: “Here we consider multimodality as a general characteristic of language, with composite utterances as instantiations of multimodality” (Janzen, 2017, p. 519). Following Enfield’s model, Janzen (2017, p. 518) defines utterance as “a complete unit of social action which always has multiple components, which is always embedded in a sequential context. . . , and whose interpretation always draws on both conventional and non-conventional signs, joined indexically as wholes (Enfield, 2009, p. 223).”

These proposals mark an important shift toward a comparative perspective between two fully fledged languages that are both multimodal. They indicate a path toward deepening and systematizing existing comparisons. One possible starting point for a systematic comparison would be to start either from a gesture studies or a sign language studies understanding of gestures. Starting from a gesture studies point of view could involve an investigation of the full spectrum of gestural forms (singular, recurrent, and emblematic gesture) in spoken and signed language use. For instance, Müller (2004) and Bressem and Müller (2014a) have documented that the palm-up-open-hand (PUOH) is widely used as a pragmatic gesture by German speakers. Steinbach has shown that it is frequently used in German Sign Language (DGS).

Starting, on the other hand, from a sign language point of view involves investigating how, for example, ‘constructed action’ (Bressem et al., 2018; Hübl and Steinbach, 2018), ‘classifiers’ (Schembri, 2003; Kendon, 2004; Müller, 2009), or the use of ‘space,’ ‘depicting verbs,’ or ‘buoys’ (Liddell, 2003) are potentially realized in co-speech gestures.

Such a comparison between two languages involves two facets: commonalities of gesture and sign resulting from a shared medium of expression (what Kendon refers to as being ‘cut from the same cloth’), and commonalities resulting from language use within and across language communities. In Germany, for instance, spoken German and German Sign Language are in close language contact (cf. Figure 7). This holds for the



community of DGS signers (the use of the PUOH documents this), but it also affects bilingual speakers of German and DGS, who may include signs in their gesturing, much as they integrate a new Anglicism into their spoken language.

In sum, this points toward dynamic relations between co-speech gestures and co-sign gestures across signed and spoken languages. The dynamic relations are motivated either by the commonality of the medium of expression or by a language-contact situation.

CONCLUSION

The systematic reconstruction of the gesture-sign relation across the history of gesture studies offered in this paper has argued that the question of how gesture and sign relate critically depends on the notion of ‘gesture’ employed. In fact, there is not one question at all, but rather a multitude of questions to be addressed. Minimally, one must separately address the question of gesture change, that is, the historical dynamics of gesture and sign, and the question of cross-linguistic comparison of spoken and signed languages. I have also shown that comparing multimodal languages in use, that involve singular, recurrent, and emblematic gestures, is different from comparing signing or speaking only with regard to singular gestures and under experimental conditions. Against this background, the ‘cataclysmic break’ diagnosed by Singleton et al. (1995) appears to be the result of the particular definition of the term ‘gesture,’ the experimental setting, and a static concept of language. Although restricting the term gesture to singular gestures makes sense in an experimental condition and to answer a specific psychologically motivated question (such as Goldin-Meadow’s focus on gestures that predict learning), it does, however, not tell us anything about other forms of gestures that we observe in speaking as well as in signing people. In fact, it excludes a broad range of gestural forms pre-hoc and thus hides that many gestures are partially or fully conventional and yet used with speech. It also makes it impossible to see that gestures differ in terms of conventionality and stabilization only gradually and not categorically. Moreover, experimental evidence based on this restrictive notion of gesture, such as the gestures speakers produce when forced to suppress speech (so-called ‘silent gesture’), is not adequate for countering linguistic observations concerning lexicalization processes that describe gesture change across time. Historical linguistics typically reconstructs processes of language change without recourse to psychological experiments. And, indeed, there is mounting evidence that, historically, not only certain lexical signs but also some grammatical ones have evolved from gesture (Wilcox and Wilcox, 1995; Janzen, 1998, 1999; Shaffer, 2000, 2002, 2004; Janzen and Shaffer, 2002; Wilcox, 2005, 2007, 2013; Wilcox et al., 2010; Shaffer and Janzen, 2016). In contrast to Goldin-Meadow and Brentari, and also in contrast to McNeill, this suggests a dynamic, continuous and ongoing process of historical change, where no cataclysmic break is involved, and no sudden rupture transforms gesture into sign from one moment to another.

Gestures produced under experimental conditions of suppressed speech cannot tell us whether speakers engaged in other discourses than narratives of visual stimuli produce gestural sequences that have more in common with ‘silent gestures’ than it appears. In fact, not much is known about gesture sequences, gesture scenarios, or local conventionalization processes occurring in naturalistic communicative contexts, but what little we know suggests that gestures are indeed often produced in complex structures involving linear as well as simultaneous productions of gestures (Müller, 2010; Müller and Tag, 2010; Müller et al., 2013a,b; Ladewig, 2014a; Bressemer et al., 2018; Ladewig, in press). A narrow focus, useful for experiments, hides the full range of gestural forms commonly employed with spoken and signed language in naturalistic contexts. From a point of view of language use, this appears as a deliberate exclusion of the scope of phenomena that fall under a ‘composite utterance’ model. If we agree, however, that language is inherently (or ‘variably,’ Cienki, 2012) multimodal, then we need cross-linguistic investigations of spoken and signed languages along the lines set out in this paper. It suggests dynamic relations across multimodal languages that are motivated by commonalities of the expressive medium and by language contact.

This brings us back to the outset of this paper and to the milestone work carried out by Adam Kendon, David McNeill, and Susan Goldin-Meadow making it unmistakably clear that the study of gestures belongs to the study of language. The controversial positions concerning the relation between gesture and sign reflect different concepts of gesture and of language. From the point of view of studying multimodal language use ‘in the wild,’ as advocated in this article, the relation between gesture and sign is to be seen as dynamic on various levels, which in turn opens up fascinating new avenues for research.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

FUNDING

The research for this project was supported by a grant from the Russian Science Foundation (project number 14-48-00067).

ACKNOWLEDGMENTS

I am indebted to Lisa Bickelmayer not only for helping with the formatting, but for many inspiring conversations on the relation between gesture and sign. I am extremely grateful to Lynne Cameron for supporting the work on this article. I thank the reviewers for their valuable critical comments on this paper and I am specifically grateful to MG for her lucid, supportive, and meticulous recommendations on this manuscript.

REFERENCES

- Behaghel, O. (1924). *Deutsche Syntax. Eine geschichtliche Darstellung*. Heidelberg: Carl Winter.
- Bellugi, U., and Newkirk, D. (1981). Formal devices for creating new signs in American sign language. *Sign Lang. Stud.* 30, 1–35. doi: 10.1353/sls.1981.0001
- Boutet, D. (2010). Structuration physiologique de la gestuelle: modèle et tests. *Lidil* 42, 77–96.
- Bressem, J., Ladewig, S. H., and Müller, C. (2013). “Linguistic annotation system for gestures (LASG),” in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1098–1124.
- Bressem, J., Ladewig, S. H., and Müller, C. (2018). “Ways of expressing action in multimodal narrations – the semiotic complexity of character viewpoint depictions,” in *Linguistic Foundations of Narration in Spoken and Sign Languages*, eds A. Hübl and M. Steinbach (Amsterdam: Benjamins), 223–249. doi: 10.1075/la.247.10bre
- Bressem, J., and Müller, C. (2014a). “A repertoire of German recurrent gestures with pragmatic functions,” in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1575–1592.
- Bressem, J., and Müller, C. (2014b). “The family of away-gestures: negation, refusal, and negative assessment,” in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1592–1604.
- Bressem, J., and Müller, C. (2017). The “negative-assessment-construction” – a multimodal pattern based on a recurrent gesture? *Linguist. Vanguard* 3, 1–9. doi: 10.1515/lingvan-2016-0053
- Brinton, L. J., and Traugott, E. C. (2005). *Lexicalization and Language Change*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511615962
- Burling, R. (1999). “Motivation, conventionalization, and arbitrariness in the origin of language,” in *The Origins of Language: What Human Primates Can Tell Us*, ed. B. J. King (Santa Fe, NM: School of American Research Press), 307–350.
- Butterworth, B., and Hadar, U. (1989). “Gesture, speech and computational stages: a reply to McNeill. *Psychol. Rev.* 96, 168–174. doi: 10.1037/0033-295X.96.1.168
- Bybee, J. (2003). “Cognitive processes in grammaticalization,” in *The New Psychology of Language 2: Cognitive and functional approaches to language structure*, ed. M. Tomasello (Mahwah, NJ: Lawrence Erlbaum), 145–167.
- Bybee, J., Perkins, R., and Pagliuca, W. (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. Chicago, IL: The University of Chicago Press.
- Cienki, A. (2012). Usage events of spoken language and the symbolic units we (may) abstract from them. *Cogn. Process. Lang.* 25, 149–158.
- Cienki, A., and Iriskhanova, O. (eds) (2018). *Aspectuality Across Languages. Event Construal in Speech and Gesture*. Amsterdam: Benjamins.
- Cienki, A., and Müller, C. (2014). “Ways of viewing metaphor in gesture,” in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1766–1781.
- Clark, H. (1996). *Using Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620539
- Clark, H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123, 324–347. doi: 10.1037/rev0000026
- Copple, M. M. (2013). “Enlightenment philosophy: gestures, language, and the origin of human understanding,” in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* 38.1, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 378–392.
- Croft, W. (2012). *Verbs: Aspect and Causal Structure*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199248582.001.0001
- Dudis, P. G. (2011). “Some observations on form-meaning correspondences in two types of verbs in ASL,” in *Deaf Around the World: The Impact of Language*, eds G. Mathur and D. J. Napoli (Oxford: Oxford University Press), 83–95.
- Efron, D. (1941). *Gesture and Environment: A Tentative Study of Some of the Spatio-temporal and Linguistic Aspects of the Gestural Behavior of Eastern Jews and Southern Italians in New York City, Living Under Similar as Well as Different Environmental Conditions*. New York, NY: Kings Crown Press.
- Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal behavior. Categories, origins, usage and coding. *Semiotica* 1, 49–98. doi: 10.1515/semi.1969.1.1.49
- Emmorey, K. (1999). “Do signers gesture?” in *Gesture, Speech, and Sign*, eds L. Messing and R. Campbell (New York, NY: Oxford University Press), 133–159. doi: 10.1093/acprof:oso/9780198524519.003.0008
- Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511576737
- Enfield, N. J. (2013). “A ‘composite utterances’ approach to meaning,” in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction* 38.1, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 689–707.
- Engle, R. A. (1998). “Not channels but composite signals: speech, gesture, diagrams, and object demonstrations are integrated in multi-modal explanations,” in *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, eds M. A. Gernsbacher and S. J. Derry (Mahwah, NJ: Lawrence Erlbaum Associates Inc.), 321–326.
- Feyereisen, P. (1987). Gestures and speech, interactions and separations: a reply to McNeill. *Psychol. Rev.* 94, 493–498. doi: 10.1037/0033-295X.94.4.493
- Fricke, E. (2013). “Towards a unified grammar of gesture and speech: a multimodal approach,” in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction* 38.1, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 733–754.
- Frishberg, N. (1975). Arbitrariness and iconicity. *Language* 51, 696–719. doi: 10.2307/412894
- Goldin-Meadow, S. (2015). The impact of time on predicate forms in the manual modality: signers, homesigners, and silent gesturers. *Topics* 7, 169–184. doi: 10.1111/tops.12119
- Goldin-Meadow, S., and Brentari, D. (2017). Gesture, sign, and language: the coming of age of sign language and gesture studies. *Behav. Brain Sci.* 40, 1–17. doi: 10.1017/S0140525X15001247
- Gullberg, M. (1998). *Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish*. Lund: Lund University Press.
- Hadar, U. (2013). “Co-verbal gestures: between communication and speech production,” in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* 38.1, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 804–821.
- Heine, B., Claudi, U., and Hünemeyer, F. (1991). *Grammaticalization: A Conceptual Framework*. Chicago, IL: University of Chicago Press.
- Heine, B., and Kuteva, T. (2007). *The Genesis of Grammar: A Reconstruction*. Oxford: Oxford University Press.
- Holt, J. (1943). *Études D’aspect*. København: Universitetsforlaget i Aarhus ejnar munksgaard, Vol. 15, 1–84. doi: 10.1017/CBO9781139165525
- Hopper, P. J., and Traugott, E. C. (2003). *Grammaticalization*, 2nd Edn. Cambridge: Cambridge University Press. doi: 10.1075/la.247
- Hübl, A., and Steinbach, M. (2018). *Linguistic Foundations of Narration in Spoken and Sign Languages*. Amsterdam: Benjamins.
- Janzen, T. (1995). *The Polygrammaticalization of FINISH in ASL*. Available at: https://mspace.lib.umanitoba.ca/bitstream/handle/1993/18855/Janzen_The_polygrammaticalization.pdf?sequence=1
- Janzen, T. (1998). *Topicality in ASL: Information ordering, Constituent Structure, and the Function of Topic Marking*. Ph.D. thesis, University of New Mexico, Albuquerque: NM. doi: 10.1075/sl.23.2.03jan
- Janzen, T. (1999). The grammaticization of topics in American Sign Language. *Stud. Lang.* 23, 271–306.
- Janzen, T. (2012). “Lexicalization and grammaticalization,” in *Sign Language: An International Handbook*, eds M. Steinbach, R. Pfau, and B. Woll (Berlin: Mouton de Gruyter), 816–841. doi: 10.1515/cog-2016-0121

- Janzen, T. (2017). Composite utterances in a signed language: topic constructions and perspective-taking in ASL. *Cogn. Linguist.* 28, 511–538.
- Janzen, T., and Shaffer, B. (2002). "Gesture as the substrate in the process of ASL grammaticization," in *Modality and Structure in Signed and Spoken Languages*, eds R. P. Meier, K. Cormier, and D. Quinto-Pozos (Cambridge: Cambridge University Press), 199–223. doi: 10.1075/msw.1.2.02kap
- Kappelhoff, H., and Müller, C. (2011). Embodied meaning construction. Multimodal metaphor and expressive movement in speech, gesture, and feature film. *Metaphor Soc. World* 1, 121–153.
- Kendon, A. (1972). "Some relationships between body motion and speech," in *Studies in Dyadic Communication*, eds A. Siegman and B. Pope (Elmsford, NY: Pergamon Press), 177–216.
- Kendon, A. (1975). Gesticulation, speech and the gesture theory of language origins. *Sign Lang. Stud.* 1975, 349–373. doi: 10.1515/semi.1980.31.1-2.1
- Kendon, A. (1980a). A description of a deaf-mute sign language from the Enga Province of Papua New Guinea with some comparative discussion. Part I: the formal properties of Enga signs. *Semiotica* 32, 1–34. doi: 10.1016/j.wneu.2017.05.170
- Kendon, A. (1980b). A description of a deaf-mute sign language, etc. Part II: the semiotic functioning of Enga signs. *Semiotica* 32, 81–117. doi: 10.1016/j.wneu.2017.05.170
- Kendon, A. (1980c). A description of a deaf-mute sign language, etc. Part III: aspects of utterance construction. *Semiotica* 32, 245–313. doi: 10.1016/j.wneu.2017.05.170
- Kendon, A. (1980d). "Gesture and speech: two aspects of the process of utterance," in *Nonverbal Communication and Language*, ed. M. R. Key (The Hague: Mouton), 207–227.
- Kendon, A. (1981). Geography of gesture. *Semiotica* 37, 129–163.
- Kendon, A. (1983). "Gesture and speech: how they interact," in *Nonverbal Interaction* (=Sage Annual Reviews of Communication, Vol. 11, eds J. M. Wiemann and R. P. Harrison (Beverly Hills, CA: Sage Publications), 13–46.
- Kendon, A. (1984). "Did gesture have the happiness to escape the curse at the confusion of Babel?," in *Nonverbal Behavior: Perspectives, Applications, Intercultural Insights*, ed. A. Wolfgang (Lewiston, NY: C.J. Hogrefe), 75–114.
- Kendon, A. (1988a). "How gestures can become like words," in *Crosscultural Perspectives in Nonverbal Communication*, ed. F. Poyatos (Toronto, CJ: Hogrefe), 131–141.
- Kendon, A. (1988b). *Sign Languages of Aboriginal Australia: Cultural, Semiotic and Communicative Perspectives*. Cambridge: Cambridge University Press.
- Kendon, A. (1992). Some recent work from Italy on quotable gestures ('emblems'). *J. Linguist. Anthropol.* 2, 77–93. doi: 10.1525/jlin.1992.2.1.92
- Kendon, A. (2000). "Language and gesture: unity or duality," in *Language and Gesture: Window into Thought and Action*, ed. D. McNeill (Cambridge: Cambridge University Press), 47–63.
- Kendon, A. (2002). "Historical observations on the relationship between research on sign languages and language origins theory," in *The Study of Signed Languages: Essays in Honor of William C. Stokoe*, eds D. Armstrong, M. A. Karchmar, and J. V. van Cleve (Washington, DC: Gallaudet University Press), 32–52.
- Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572
- Kendon, A. (2008). Some reflections on the relationship between 'gesture' and 'sign'. *Gesture* 8, 348–366. doi: 10.1075/gest.8.3.05ken
- Kendon, A. (2013). "Exploring the utterance roles of visible bodily action: a personal account," in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction 38.1*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 7–28.
- Kendon, A. (2014). Semiotic diversity in utterance production and the concept of 'language'. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130293. doi: 10.1098/rstb.2013.0293
- Klima, E. S., and Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.
- Kyle, J. G., and Woll, B. (1985). *Sign Language. The Study of Deaf People and their Language*. Cambridge: Cambridge University Press.
- Ladewig, S. H. (2010). Beschreiben, suchen und auffordern –varianten einer rekurrenten geste. *Zeitschrift für Sprache und Literatur* 41, 89–111.
- Ladewig, S. H. (2011). Putting the cyclic gesture on a cognitive basis. *CogniTextes* 6. Available at: <http://cognitextes.revues.org/406>
- Ladewig, S. H. (2014a). "Creating multimodal utterances. The linear integration of gestures into speech," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction 38.2*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 1662–1677.
- Ladewig, S. H. (2014b). "Recurrent gestures," in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction 38.2*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 1558–1575.
- Ladewig, S. H. (2014c). "The cyclic gesture," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction 38.2*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 1605–1618.
- Ladewig, S. H. (in press). *Integrating Gestures – Cognitive Grammar Multimodal*. Berlin, NY: De Gruyter Mouton.
- Lane, H. (1979). *The Wild Boy of Aveyron*. Cambridge, MA: Harvard University Press.
- Liddell, S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511615054
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychol. Rev.* 92, 350–371.
- McNeill, D. (1986). Iconic gestures of children and adults. *Semiotica* 62, 107–128. doi: 10.1515/semi.1986.62.1-2.107
- McNeill, D. (1987). So you do think gestures are nonverbal! Reply to Feyereisen. *Psychol. Rev.* 94, 499–504. doi: 10.1037/0033-295X.94.4.499
- McNeill, D. (1989). A straight path – to where? Reply to butterworth and hadar. *Psychol. Rev.* 96, 175–179. doi: 10.1037/0033-295X.96.1.175
- McNeill, D. (1992). *Hand and Mind. What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.
- McNeill, D. (2000). "Introduction" in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 1–10. doi: 10.1017/CBO9780511620850
- McNeill, D. (2012). *How Language Began. Gesture and Speech in Human Evolution*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139108669
- McNeill, D. (2013). "The co-evolution of gesture and speech, and downstream consequences," in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction 38.1*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 480–512.
- McNeill, D., and Sowa, C. (2011). "The birth of a morph," in *Integrating Gestures: The Interdisciplinary Nature of Gesture*, eds G. Stam and M. Ishino (Amsterdam: Benjamins), 27–47.
- Mittelberg, I. (2013). "The embodied mind: Cognitive-semiotic principles as motivating forces in gesture," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction 38.1*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and S. Teßendorf (Berlin, NY: Mouton de Gruyter), 750–779.
- Mittelberg, I. (2014). "Gestures and iconicity," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction 38.2*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 1712–1732.
- Mittelberg, I., and Evola, V. (2014). "Iconic and representational gestures," in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction 38.2*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 1732–1746.
- Mittelberg, I., and Waugh, L. R. (2014). "Gestures and metonymy," in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction 38.2*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 1747–1766.

- Müller, C. (1998a). "Iconicity and gesture," in *Oralité et Gestualité: Communication Multimodale, Interaction*, ed. S. Santi (Montréal, QC: L'Harmattan), 321–328.
- Müller, C. (1998b). *Rede begleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Berlin Verlag Arno Spitz.
- Müller, C. (2000). "Zeit als raum. Eine kognitiv-semantische mikroanalyse des sprachlichen und gestischen ausdrucks von aktionsarten," in *Botschaften Verstehen. Kommunikationstheorie und Zeichenpraxis. Festschrift für Helmut Richter*, eds E. W. B. Hess-Lüttich and H. Walter Schmitz (Frankfurt: Peter Lang), 211–228.
- Müller, C. (2004). "The palm-up-open-hand. A case of a gesture family?," in *The Semantics and Pragmatics of Everyday Gestures. The Berlin Conference*, eds C. Müller and R. Posner (Berlin: Weidler), 233–256.
- Müller, C. (2007a). "A dynamic view on metaphor, gesture and thought," in *Gesture and the Dynamic Dimension of Language. Essays in Honor of David McNeill*, eds S. D. Duncan, J. Cassell, and E. T. Levy (Amsterdam: John Benjamins), 109–116. doi: 10.1075/gs.1.12mul
- Müller, C. (2007b). *A Semiotic Profile: Adam Kendon. Semiotix*, Vol. 9. Available at: <https://semioticon.com/sx-old-issues/semiotix9/sem-9-03.html>
- Müller, C. (2008). *Metaphors Dead and Alive, Sleeping and Waking. A Dynamic View*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226548265.001.0001
- Müller, C. (2009). "Gesture and language," in *Routledge's Linguistics Encyclopedia*, ed. K. Malmkjaer (Abington: Routledge), 214–217.
- Müller, C. (2010). Wie gesten bedeuten. eine kognitiv-linguistische und sequenzanalytische perspektive. *Sprache und Literatur* 41, 37–68.
- Müller, C. (2014a). "Gestures as 'deliberate expressive movement'," in *From Gesture in Conversation to Visible Action as Utterance*, eds M. Seyfeddinipur and M. Gullberg (Amsterdam: John Benjamins), 127–152.
- Müller, C. (2014b). "Gestural Modes of Representation as techniques of depiction," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 1687–1702.
- Müller, C. (2014c). "The ring across space and time: variation and stability of forms and meanings," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 1511–1522.
- Müller, C. (2016). "From mimesis to meaning: a systematics of gestural mimesis for concrete and abstract referential gestures," in *Meaning, Mind and Communication: Explorations in Cognitive Semiotics*, eds J. Zlatev, G. Sonesson, and P. Konderak (Frankfurt am Main: Peter Lang), 211–226.
- Müller, C. (2017). How recurrent gestures mean: conventionalized contexts-of-use and embodied motivation. *Gesture* 16, 278–306.
- Müller, C. (2018). "Gestures as motion events," in *Aspectuality Across Languages. Event Construal in Speech and Gesture*, eds A. Cienki and O. Irishkanova (Amsterdam: Benjamins).
- Müller, C., Bressemer, J., and Ladewig, S. H. (2013a). "Towards a grammar of gesture—a form based view," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* 38.1, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 707–733.
- Müller, C., Ladewig, S. H., and Bressemer, J. (2013b). "Gesture and speech from a linguistic point of view – a new field and its history," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* 38.1, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 55–81.
- Müller, C., and Kappelhoff, H. (2018). "Cinematic metaphor: experience – affectivity – temporality," in *coll. with*, eds S. Greifenstein, D. Horst, and C. Schmitt (Berlin: De Gruyter/Mouton).
- Müller, C., and Tag, S. (2010). The dynamics of metaphor: foregrounding and activating metaphoricity in conversational interaction. *Cogn. Semiotics* 6, 85–120. doi: 10.3726/81610_85
- Neumann, R. (2004). "The conventionalization of the German ring gesture in German discourse," in *The Semantics and Pragmatics of Everyday Gestures*, eds C. Müller and R. Posner (Berlin: Weidler), 217–224.
- Parrill, F. (2008). "Form, meaning, and convention: a comparison of a metaphoric gesture with an emblem," in *Metaphor and Gesture*, eds A. Cienki and C. Müller (Amsterdam: John Benjamins), 195–217.
- Pfau, R., and Steinbach, M. (2011). "Grammaticalization in Sign Languages," in *The Oxford Handbook of Grammaticalization*, eds H. Narrog and B. Heine (Oxford: Oxford University Press), 683–695.
- Schembri, A. (2003). "Rethinking 'classifiers' in signed languages," in *Perspectives on Classifier Constructions in Sign Languages*, ed. K. Emmorey (Mahwah, NJ: Lawrence Erlbaum), 3–34. doi: 10.1353/sls.1981.0011
- Scroggs, C. L. (1981). The use of gesturing and pantomiming: the language of a nine year old deaf boy. *Sign Lang. Stud.* 30, 61–77.
- Shaffer, B. (2000). *A Syntactic, Pragmatic Analysis of the Expression of Necessity and Possibility in American Sign Language*. Ph.D. thesis, Albuquerque, NM, University of New Mexico. doi: 10.1353/sls.2002.0026
- Shaffer, B. (2002). CAN'T: the negation of modal notions in ASL. *Sign Lang. Stud.* 3, 34–53. doi: 10.1515/cogl.2004.007
- Shaffer, B. (2004). Information ordering and speaker subjectivity: modality in ASL. *Cogn. Linguist.* 15, 175–195.
- Shaffer, B., and Janzen, T. (2016). "Modality and mood in American Sign Language," in *The Oxford Handbook of Modality and Mood*, eds J. Nuyts and J. van der Auwera (Oxford: Oxford University Press), 448–469.
- Singleton, J. L., Goldin-Meadow, S., and McNeill, D. (1995). "The cataclysmic break between gesticulation and sign: evidence against an evolutionary continuum of manual communication," in *Language, Gesture, and Space*, eds K. Emmorey and J. Reilly (Hillsdale, NJ: Erlbaum Associates), 287–311.
- Slama-Cazacu, T. (1976). "Nonverbal components in message sequence: 'Mixed syntax,'" in *Language and Man: Anthropological Issues*, eds W. C. McCormack and S. A. Wurm (The Hague: Mouton), 217–227.
- Steinbach, M., Pfau, R., and Woll, B. (2012). *Sign Language: An International Handbook*. Berlin & NY: Mouton de Gruyter.
- Streeck, J. (2009). *Gesturecraft: The Manu-Facture of Meaning*. Amsterdam: John Benjamins. doi: 10.1075/gs.2
- Streeck, J. (2013). "Praxeology of gesture," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* 38.1, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 674–688.
- Streeck, J. (2017). *Self-Making Man. A Day of Action, Life, and Language*. New York, NY: Cambridge University Press. doi: 10.1017/9781139149341
- Supalla, T. (1982). *Structure and acquisition of verbs of motion and location in American Sign Language*. Ph.D. thesis, San Diego, University of California.
- Supalla, T. (1986). "The classifier system in American Sign Language," in *Noun Classification and Categorization*, ed. C. Craig (Amsterdam: John Benjamins), 181–214.
- Teßendorf, S. (2013). "Emblems, quotable gestures, or conventionalized body movements," in *Body – Language – Communication: An international Handbook on Multimodality in Human Interaction* 38.1, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 82–100.
- Van Loon, E., Pfau, R., and Steinbach, M. (2014). "The grammaticalization of gestures in sign languages," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 2133–2149.
- Vermeerbergen, M., and Demey, E. (2007). "Sign + gesture = speech + gesture? Comparing aspects of simultaneity in Flemish Sign Language to instances of concurrent speech and gesture," in *Simultaneity in Signed Languages*, eds M. Vermeerbergen, L. Leeson, and O. Crasborn (Amsterdam: John Benjamins), 257–282.
- Wilcox, P. (1998). "GIVE: acts of giving in American Sign Language," in *The Linguistics of Giving*, ed. J. Newman (Amsterdam: John Benjamins), 175–207. doi: 10.1075/tsl.36.07wil
- Wilcox, S. (2005). Routes from gesture to language. *Revista da abralin* 4, 11–45.
- Wilcox, S. (2007). "Routes from gesture to language," in *Verbal and Signed Languages: Comparing Structures, Constructs and Methodologies*, eds E. Pizzuto, P. Pietrandrea, and R. Simone (Berlin: Mouton de Gruyter), 107–131.
- Wilcox, S. (2013). "Speech, sign and gesture," in *Body – Language – Communication: An International Handbook on Multimodality in Human*

- Interaction 38.1*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton).
- Wilcox, S., Rossini, P., and Pizzuto, E. (2010). "Grammaticalization in sign languages," in *Sign Languages*, ed. D. Brentari (Cambridge: Cambridge University Press), 332–354. doi: 10.1017/CBO9780511712203.016
- Wilcox, S., and Wilcox, P. (1995). "The gestural expression of modality in ASL," in *Modality in Grammar and Discourse*, eds J. Bybee and S. Fleischman (Amsterdam: John Benjamins), 135–162. doi: 10.1075/tsl.32.07wil

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Müller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Body as Evidence for the Nature of Language

Wendy Sandler*

Sign Language Research Laboratory, University of Haifa, Haifa, Israel

Taking its cue from sign languages, this paper proposes that the recruitment and composition of body actions provide evidence for key properties of language and its emergence. Adopting the view that compositionality is the fundamental organizing property of language, we show first that actions of the hands, face, head, and torso in sign languages directly reflect linguistic components, and illuminate certain aspects of compositional organization among them that are relevant for all languages, signed and spoken. Studies of emerging sign languages strengthen the approach by showing that the gradual recruitment of bodily articulators for linguistic functions directly maps the way in which a new language increases in complexity and efficiency over time. While compositional communication is almost exclusively restricted to humans, it is not restricted to language. In the spontaneous, intense emotional displays of athletes, different emotional states are correlated with actions of particular face and body features and feature groupings. These findings indicate a much more ancient communicative compositional capacity, and support a paradigm that includes visible body actions in the quest for core linguistic properties and their origins.

OPEN ACCESS

Edited by:

Manuel Carreiras,
Basque Center on Cognition, Brain
and Language, Spain

Reviewed by:

Michael Charles Corballis,
University of Auckland, New Zealand
Brendan Costello,
Basque Center on Cognition, Brain
and Language, Spain

*Correspondence:

Wendy Sandler
wendy.sandler@gmail.com

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 09 February 2018

Accepted: 03 September 2018

Published: 29 October 2018

Citation:

Sandler W (2018) The Body as
Evidence for the Nature of Language.
Front. Psychol. 9:1782.
doi: 10.3389/fpsyg.2018.01782

Keywords: sign language, compositionality, embodiment, language emergence, language evolution, emotion

INTRODUCTION

Sign languages and spoken languages differ dramatically in the physical modality of transmission. Despite this difference, since sign languages have been taken seriously as full natural languages, investigators have placed the emphasis on the numerous similarities between the two systems. In *Sign Language and Linguistic Universals* (Sandler and Lillo-Martin, 2006), all chapters but one adopt theories devised on the basis of spoken language to analyze the morphology, phonology, prosody, and syntax of sign languages. Though the physical manifestations of linguistic properties are duly described, the research paradigm works from linguistic theory to its manifestation by the body – from the linguistic mechanisms in the mind out to the body. Only the final chapter of the book deals with so-called modality effects that distinguish the form of sign language from that of spoken language. Here the direction of investigation is reversed. Working from the body to language, from the outside in, I bring together a range of diverse studies to show that the recruitment and composition of body actions provide direct evidence for linguistic properties and their emergence.

Since the beginning of linguistics, the main object of study has been the structure and arrangement of words. This focus has been attributed to the technology of writing, which made it possible to record these parts of language, so that they could be studied scientifically (Downing et al., 1992). As a result, the elements that can be recorded in writing, the principles behind them, and the meanings associated with them, became the primary data. Among the effects of writing systems is the segmental view of the language signal as beads on a string

(see e.g., De Saussure, 1959; Aronoff, 1992). The technology of writing systems facilitated linguistic analysis, and linguists have made much progress over the last century and more using words and their arrangement as data.

Writing has undoubtedly advanced civilization, but it is not a component of the language faculty. According to *Ethnologue* (Simons and Fenning, 2018), fewer than half of the world's languages have writing systems, and for most of those that do, there are large populations of speakers who are illiterate in the language, and have not achieved what Gough and Hillinger (1980) famously called 'an unnatural act' – learning to read. Furthermore, standard written languages like Chinese, English, or Hindi, almost never represent a person's actual spoken language. The human language capacity is independent of the written word. The fact that it is possible to convey much of the (spoken) language message in writing is of interest, but it is also deceptive.¹

The language faculty is intimately entrenched in the body – not only in the voice, but in the face, the hands, and the torso as well. In recent decades, technology for recording language has advanced greatly, and it is now easy to capture and study both the auditory and visual signals that are the physical substance of language – what we actually produce and perceive. These advances have influenced the study of phonetics, phonology, and intonation, fostering new approaches such as articulatory phonology (Browman and Goldstein, 1992). Through video technology, we can now observe gestures and facial expressions, facilitating the much younger but thriving field of co-speech gesture (McNeill, 1992; Kendon, 2004; Müller et al., 2013, 2014; Church et al., 2017). These technological advances allow us to study the interaction between the auditory and visual domains in spoken language. By including visually perceived bodily signals in our understanding of human language, we put language back in the body, and humans in their ecological evolutionary setting.²

In the natural and spontaneous languages of deaf communities, there is no language at all without the visible bodily signal. Technological advances have also made it possible to study these languages rigorously; for example, the early and seminal research of Klima and Bellugi (1979) relied partly on videotaped data. Sign languages emerge spontaneously and relatively quickly whenever deaf people have an opportunity to communicate regularly (e.g., Senghas, 2003; Sandler et al., 2005), and even individual deaf children in hearing, speaking households create gestural systems with the seeds of linguistic structure (Goldin-Meadow, 2003). It is now accepted that sign languages are a manifestation of a universal human linguistic

endowment. It follows that they should not be regarded as extraneous or peripheral, but rather as fundamental to our understanding of language.

Taking its cue from sign languages, this article pulls together results from a range of studies to support the proposal that the recruitment and composition of body actions count as primary evidence for linguistic properties and their emergence. This approach has two aims. The first relates to sign language and co-speech gesture; and the second relates to all language. The first aim is to motivate a model of the relation between linguistic functions and bodily actions in sign languages, and a principled way of relating that model to co-speech gesture. The second is to give the human body a focal role in the pursuit of knowledge about core properties of language, how they interact, how they emerge in new languages, and how they evolved. The approach complements and supplements those that study only mind-internal computational manipulations that create language structure (see e.g., Chomsky, 2007).³

A single thread that unifies all modern linguistic research is that the human language capacity is rooted in our ability to communicate compositionally. Compositionality was first introduced by Frege (1914/1979) as a constraint on the relation between syntax and semantics (see Hinzen et al., 2012). This versatile capacity is a robust human trait. Other species, such as non-human primates, can certainly command compositionally organized cognitive operations and social systems, which may indeed have provided primordial underpinning for compositional expression (see the section on Language Evolution below). However, to date, evidence for compositionality in the communicative capacity of other species is scant.⁴ The version of the compositionality principle assumed here is given in (1).

- (1) The compositionality principle (Szabó, 2012, p. 71).

The meaning of a complex expression is determined by the meanings its constituents have individually and the way those constituents are combined.

Complex words can be understood in terms of their component parts, and the same is true of phrases, clauses, complex sentences, and so forth. It is understood that each component can be recombined with other components, within the constraints of the system, to create new complex forms.

³Exceptions to the tendency to ignore the body are the disciplines of phonology and intonation, which commonly attribute universal generalizations to the nature of the articulatory and perceptual systems and their transmission and acquisition (e.g., Browman and Goldstein, 1992; Archangeli and Pulleyblank, 1996; Blevins, 2004, 2012; Gussenhoven, 2004).

⁴Other species, notably, birds, have combinatorial structure in their communication – elements combine and recombine – but the components and their recombinations are usually not considered meaningful (e.g., Wohlgemuth et al., 2010). There is some literature demonstrating limited compositionality based on laboratory experiments manipulating tones in birdsong (Suzuki et al., 2017). Non-human primates communicate multi-modally (Liebal et al., 2013), but the authors do not present evidence of compositionality. Arnold and Zuberbühler (2012) write that two-part meaningful vocal components recombine in non-human primates, and a single vocal signal in putty-nosed monkeys appears to modify meaning in combination with other components (Schlenker et al., 2016). To my knowledge, no evidence has been presented of complex combinations with reliable interpretations in non-human species.

¹A reviewer pointed out that language can be communicated without the visual component – for example, in telephone conversations. However, people use manual and facial gestures, sometimes prolifically, in phone conversations as well, and both the degree to which telephone speakers produce added linguistic information as compensation, and the degree to which information is lost to the perceiver in these situations, have yet to be studied rigorously. Direct deictic expressions such as 'that' and 'there' must be accompanied by gesture, and are infelicitous in telephone conversations. That bodily gesture is a basic component of linguistic communication is attested by the fact that congenitally blind people also gesture when they speak (Iverson and Goldin-Meadow, 1998).

²Throughout, reference to humans refers only to contemporary homo sapiens and not to any predecessors.

Though compositionality does not exhaustively account for all of language structure, the basic principle is robust and results in productivity and creativity in the language of humans, and of humans alone.

In what follows, motivation for the body-as-evidence approach, in which the body and compositionality figure prominently, comes from four directions: (1) established sign languages, (2) language emergence (of which the only empirical data are from sign languages), (3) gesture, and (4) communicative displays of intense emotion in a human compositional system that is far more ancient than language.

The idea that sign languages are fully fledged linguistic systems at all levels of structure is by now widely accepted across the scientific community (Stokoe, 1960; Klima and Bellugi, 1979; Sandler and Lillo-Martin, 2006; Pfau et al., 2012). But inadvertently, indirectly, and somewhat myopically, the written word and the language-as-computation paradigm have dominated sign language research, as they have that of spoken language.

Sign languages, by their very nature, convey linguistic information directly through articulations of different parts of the body – an advantage for linguistic analysis that is typically overlooked. It can be no accident that (apart from differences in detail of the kind that any grammatical system would exhibit, due to conventionalization and automaticity) unrelated sign languages tend to achieve this kind of structuring in very similar ways. The section on Established Sign Languages demonstrates that what I call the Grammar of the Body, which reflects universal elements of meaning and structure in a way that speech cannot. The role of iconicity in this system, all the way down to and including the phonology, is addressed, and considered in light of recent demonstrations of iconicity in spoken language.

Another advantage offered by sign languages is their youth. It is only in sign languages that language emergence, the topic of the section on The Composition of Language Emergence can be observed empirically, since it is only these languages that can emerge *de novo* at any time. In the initial stages of language emergence, we do not see the sophisticated associations between body and language form found in established sign languages. Research on Al-Sayyid Bedouin Sign Language (ABSL), an emerging language in a Bedouin village, summarized in its own section below, suggests that the gradual recruitment of parts of the body, as well as the refinement of these articulations and their interactions over time, reveals the way in which linguistic organization emerges, step by step (Sandler, 2012a). The body-based approach advocated here reveals which components of language organization arise earlier than others. We infer that these early components are critical for successful linguistic interaction. The next section then summarizes support for this perspective from another young sign language that arose under different social and linguistic conditions, Israeli Sign Language. Broadly speaking, sign languages tend to have similar body-to-language representations, suggesting that they derive from a universal, gestural base common to all of us.

The section, Gesture briefly cites related observations from the field of co-speech gesture studies. The goal is to show how the Grammar of the Body found in sign languages is tapped

by gesture as well, supporting the view that gesture provides a universal base for the systematic and constrained system underlying sign languages.

Since compositional communication is very limited or non-existent in other species (see footnote 4), but robust in humans, the question of its evolutionary origins is of interest. Some comments about different views of language evolution open the section that probes The Roots of Compositional Expression in Intense Emotional Displays. A search for the foundations of bodily compositionality leads to the study of body signals in humans that are communicative but non-linguistic, and that have internal compositional organization: body displays of intense emotion. We review our recent experiments, which analyze displays of winning and losing athletes (Cavichio et al., 2018). Interpretation of these displays – minutely coded for features of face and body – form the basis of a compositional model of the expression of emotion, illustrated for the first time here by idealized computer-generated 3-D images. This evidence from the body suggests ancient roots for compositional communication in humans.

The final section brings together these strands of research, to offer a basis for incorporating the body into future investigations of the nature of language.

BODY AND LANGUAGE STRUCTURE IN ESTABLISHED SIGN LANGUAGES

One of the most important differences between signed and spoken language is that, in sign language alone, movements of articulators (of the face, hands, and body) correspond directly to specific linguistic functions. This situation is quite unlike speech, in which movements of the vocal apparatus in themselves typically do not signify linguistic categories directly. That is, the relation between linguistic form and movement of any part of the vocal tract and the resulting acoustic signal is indirect. Across sign languages, despite expected grammatical differences, the same fundamental correspondences between bodily actions and types of linguistic functions seem to hold. This strongly suggests that sign languages are tapping deeper body-meaning correspondences, common to us all, and converting them into rule-governed linguistic systems.

The correspondences are identifiable and reveal compositional structure inherent in signed words themselves as well as in the organization of sign languages at higher levels. The following sections look selectively at the linguistic roles played by the hands, the face, the torso, and the non-dominant hand independently. The evidence points to a deeper source: the relation between communicative conceptualization and the body – for all of us.

The rich cross-linguistic literature on spoken languages is unfortunately not paralleled in the relatively young field of sign language research. In the discussion that follows, data and analyses are presented from several, often unrelated sign languages. Unless otherwise stipulated, the general characteristics described below are, to the best of my knowledge,

representative of sign languages in general. Broader cross-linguistic confirmation and grammatical detail await future empirical research.

The Hands: Iconicity and Dual Duality of Patterning

In all sign languages, the hand or the two hands together produce forms equivalent to words. Contrary to popular belief that preceded scientific sign language research (e.g., Bloomfield, 1933), Stokoe (1960) demonstrated conclusively that signs are not holistic gestures. They are composed of units of handshape, location, and movement, which make contrasts and in other ways function like the meaningless phonemes and features of spoken language. For example, **Figure 1** shows minimal pairs in Israeli Sign Language, distinguished only by features of handshape (1a), location (1b), and movement (1c).

This means that sign languages share with spoken languages the design feature named “duality of patterning” by Hockett (1960); called ‘double articulation’ by Martinet (1960): words in both modalities are comprised of both meaningless (phonological) and meaningful levels of structure. Stokoe’s non-trivial claim has been further investigated, corroborated, and refined by other researchers (e.g., Liddell and Johnson, 1986; Sandler, 1989, 2012b, 2017; van der Hulst, 1993; Brentari, 1998). The handshape, location, and movement units behave like meaningless phonological elements in the sense that their combination is constrained by their form, and they are permuted by typical phonological processes such as assimilation and deletion, which are also oblivious to meaning, targeting and influencing articulatory properties of the elements.

Evidence for a meaningless level of structure is seen in American Sign Language lexicalized compounds, which undergo the standard phonological processes of reduction and assimilation (Liddell and Johnson, 1986; Sandler, 1987, 1989, 2017). The reduction involves deletion of locations and regressive assimilation that affects the shape and orientation of the hand. The resulting compound assumes the optimal form of the prosodic word in ASL: the monosyllable (Sandler, 1999a). What is important here is that the reduction and assimilation processes affect sublexical components because of their form, irrespective of

meaning, and in fact often obscure the meaning of the individual members of the compound.

However, in their enthusiasm to demonstrate that sign languages are full languages like spoken languages, researchers often miss generalizations that result from the iconicity that is still present in the formational units of signs. That is, even as the composition and behavior of formational elements in the system tap their form regardless of meaning, the elements themselves can still bear meaning.

Iconicity goes beyond the general impression of the whole sign. A growing body of work has been describing iconic aspects of the sublexical structure of signed words (e.g., Johnston and Schembri, 1999; Fernald and Napoli, 2000; Taub, 2001; Meir, 2002, 2010; Perniss et al., 2010; Padden et al., 2013). We can say that duality of patterning in sign languages is itself double-sided: the elements that are analogous to the meaningless ‘phonemic’ units of spoken language are also often meaningful. Here sign languages and spoken languages depart, because of the iconic opportunities that the manual-visual medium so richly supports.

The semantic composition of words in any language is quite complex, even when the form is morphologically simple (Wunderlich, 2012). For example, Jackendoff (1990) analyzes the concept ‘drink’ as shown in example (2).

(2) Lexical conceptual structure of the word *drink* (Jackendoff, 1990)

drink: [event CAUSE ([thing]_i, [event GO (thing LIQUID)]_j, [path TO ([place IN ([thing MOUTH OF ([thing]_i))]])]]]

This internal structure is rarely observable in the form of the spoken word itself, e.g., *drink* in English, [ʃote] in Hebrew, *boit* in French. In any sign language, elements of the internal structure are often reflected directly, and, together, make up the meaning of a sign. Consider the sign DRINK (water) in the emerging sign language of Al-Sayyid, in **Figure 2**.

The curved hand is a container; the motion reflects causing a substance to go into something; and the mouth as place of articulation is the ‘something,’ the destination. The fact that the ‘something’ is liquid is reflected in the shape of the hand and its orientation with respect to the location, the mouth. Such signs are not pantomimes, but conventionalized signs

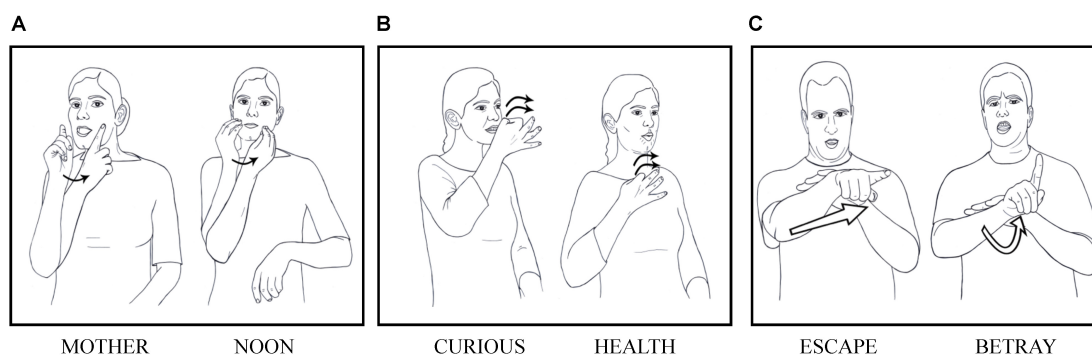


FIGURE 1 | Minimal pairs in Israeli Sign Language. Form left to right: **(Ai)** MOTHER, **(Aii)** NOON distinguished by handshape; **(Bi)** CURIOUS, **(Bii)** HEALTH distinguished by location; and **(Ci)** ESCAPE, **(Cii)** BETRAY distinguished by movement.

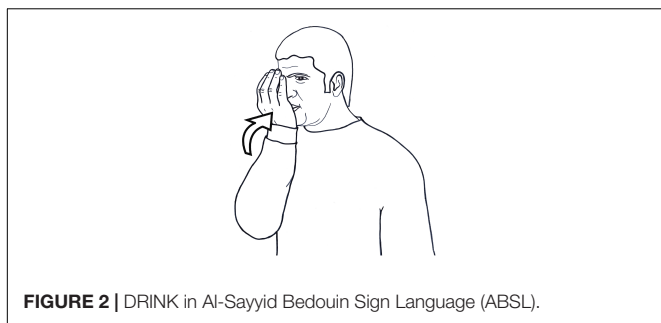


FIGURE 2 | DRINK in Al-Sayyid Bedouin Sign Language (ABSL).

and specific to ABSL. In ISL, for example, the handshape for DRINK is different from that of ABSL, derived from holding a vessel, while the movement and location still directly reflect the event of moving something liquid into the mouth location.

The sublexical components of handshape, location, and movement, which, as we see here, often retain meaning, combine according to morpho-phonological constraints (rules). These components are not morphemes in the traditional sense, since they do not serve as roots, stems, inflections or derivational elements (but see Lepic, 2015 for a different view). Nevertheless, the components are often motivated, revealing internal semantic structure, and so may be thought of as meaningful phonological elements (see Taub, 2001 for a model of meaningful sign components in ASL). Section “Iconicity in Two-Handed Signs” gives an example of iconicity of phonological elements in two-handed signs, and “Iconicity in Location and Movement” considers the phenomenon in light of recent work on iconicity in spoken language.

Iconicity in Two-Handed Signs

Recent comparative work on two-handed signs illustrates the direct relation between the internal semantic structure of a word and its bodily representation. About half of the signs in any sign language are produced with one hand; the other half are two-handed. Previous work on the

phonological structure of two-handed signs from different theoretical perspectives have often ignored or downplayed meaning (Battison, 1978; Sandler, 1989, 1993; Van der Hulst, 1996; Crasborn, 2011).

But the selection of two hands rather than one, and of the type of two-handed sign, is often motivated. Comparing lexicons of three unrelated sign languages, we have shown that signs denoting meanings that are essentially plural tend to be two-handed, more than twice as often as chance would predict (Lepic et al., 2016). Specifically, plurality, expressed in relations of composition, interaction, dimension, and relative location among entities or parts of entities, tend to be two-handed in American, Swedish, and Israeli sign languages. A subset of these signs was elicited in Al. Sayyid Bedouin Sign Language, and the results were compatible with findings for the other three sign languages. In these signs, each hand and the interaction between the two represents a component, directly revealing the composition of the concept.

For example, the sign EMPTY (Figure 3) in American, Swedish, and Israeli sign languages is unbalanced, that is, non-symmetrical, in all cases. The non-dominant hand represents a surface or container, and the dominant hand signifies its empty or unencumbered state by the type of motion it articulates in relation to the container. The two elements – an object and its empty state – are not equal in EMPTY; it is the empty state that is the salient meaning component in the concept and not the object itself. Only the dominant hand moves to signify emptiness with respect to the non-dominant hand, which signifies the surface or container, and the sign is two-handed and unbalanced in three unrelated sign languages. Enfield (2004) documented similar though unsystematic and gestural use of the two hands in the description of fish traps in Lao.

Here is the crux: particular elements of the cross-linguistic, compositional meanings of concepts, usually not overtly present in the form of spoken words, are often directly revealed in sign languages in similar ways – by the body.

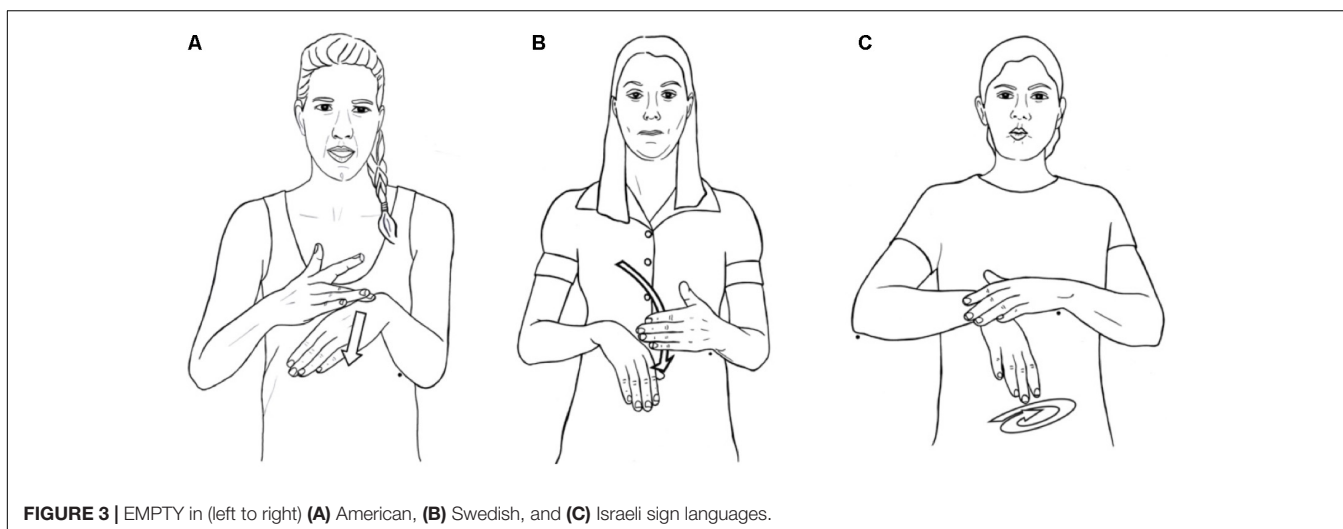


FIGURE 3 | EMPTY in (left to right) (A) American, (B) Swedish, and (C) Israeli sign languages.

Iconicity in Location and Movement

In most of the examples above, the location category of a sign can also be motivated (Fernald and Napoli, 2000; van der Kooij, 2002). For example, thought processes are typically signed on or near the upper part of the head. Movement patterns are motivated in many signs as well. **Figure 4** reveals iconicity in the movement patterns produced by the hand/s: the reciprocal, ongoing activity of negotiating motivates repeated, alternating movement of the two hands. Wilbur (2008), proposes that event structure is directly revealed in the movement pattern of verbs across sign languages. Strickland et al. (2015) provide experimental perceptual evidence from signers across sign languages and non-signers regarding movement and telicity. They write that their results “are highly suggestive that signers and non-signers share universally accessible notions of telicity as well as universally accessible “mapping biases” between telicity and visual form. (2015, p. 1).

A caveat: Not all signs are transparently iconic: many signs are arbitrary in form, and even those with iconic elements are not usually transparent – their meaning cannot be guessed by naïve observers (Klima and Bellugi, 1979). As a language matures, iconicity may diminish and signs may become more arbitrary with respect to their meaning and more constrained in form (Frishberg, 1975 for ASL; Meir and Sandler, 2008 for ISL). Furthermore, not every aspect of lexical conceptual structure is expressed iconically. For example, in the case of DRINK (example 2 and **Figure 2** above), the liquid property of what is ingested is only pragmatically inferable.⁵ Different sign languages do not always select the same meaning components for iconic representation. In addition to semantic composition, culture plays a role. If all sign languages selected the same meaning components to represent, there would be only one sign lexicon, rather than hundreds.

Yet meaning is pervasive even in formational units that behave like meaningless phonological elements; it accounts robustly for productive aspects of vocabulary formation and for similarities across sign languages. Words of sign languages, to a much greater

extent than those of spoken languages, exhibit what we might call ‘dual duality of patterning,’ and their study across sign languages will have much to reveal about the semantic composition of lexical concepts in human language generally.

Iconicity in Spoken Language

Contrary to traditional beliefs about the arbitrary relation between form and meaning in spoken language (De Saussure, 1959), instances of lexical and sublexical form have been found to have an iconic relationship (Blevins, 2012; Dingemanse et al., 2015). Blasi et al. (2016) show that some non-arbitrary associations between form and meaning are even shared across linguistic lineages, suggesting that they are not spread through language contact, but are more basic, and might even have provided an evolutionary base for language tens of millennia ago.⁶ However, the amount of iconicity in sign languages is far greater than in contemporary spoken languages, for two reasons: (1) sign languages, expressed with two visible, anatomically identical articulators, so readily avail themselves of the complex iconic representation necessary for a large vocabulary, and (2) sign languages are very young compared to spoken languages – none of them traceable farther back than 300 years (Kyle et al., 1988). Presumably, a large pool of arbitrary signal-form relations requires time to develop.⁷

The recent investigations into iconicity in spoken language in fact only serve to reinforce the claim that iconicity in sign languages can reveal universal properties of language that are not – or are no longer – as prevalent in spoken languages. Sign languages teach us that meaninglessness in duality of patterning of human language lies on a continuum and is not absolute.

The Face

In sign language after sign language, particular aspects of information structure are signaled by the upper face – brows and eyes – and by head position on the front/back axis.⁸ Across sign languages, raised brows and often head forward accompany yes-no (polar) questions, while furrowed brows accompany wh- (content) questions (ASL, Liddell, 1980; Sign language of the Netherlands, Coerts, 1992; British SL, Sutton-Spence and Woll, 1999; other sign languages, Zeshan, 2004). Squinted eyes reliably accompany shared (but not highly accessible) information, another information structuring device in ISL (Dachkovsky and Sandler, 2009; Sandler et al., accepted), and have been observed for the same function in ASL (Dachkovsky et al., 2013) and Danish Sign Language (Engberg-Pedersen, 1993), three unrelated languages.

⁵I thank a reviewer for noticing this.

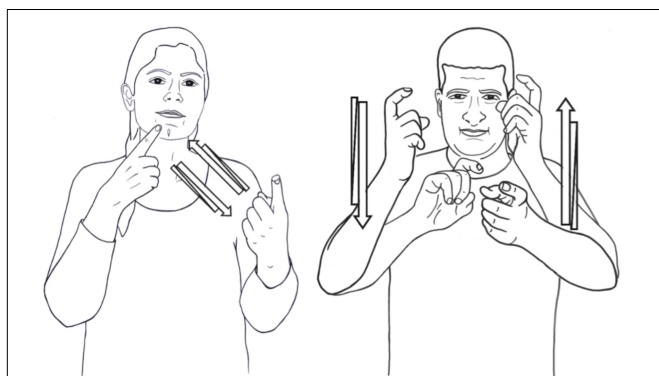


FIGURE 4 | Iconic movement in signs meaning NEGOTIATE in ISL and ABSL.

⁶I thank a reviewer for bringing this article to my attention.

⁷Frishberg (1975) shows that signs can become more arbitrary and less iconic over time, and Aronoff et al. (2005) show that iconic types of morphological complexity are common across sign languages, while arbitrary derivational morphology (often the result of grammaticalization in spoken languages) is much more rare in these young languages.

⁸In this discussion, we do not deal with affective or emotional facial expressions.

In our work, we confirm on functional and distributional grounds the earlier suggestion that facial expressions comprise the intonational component of prosody in sign languages (Reilly et al., 1990), and demonstrate that these signals are compositionally organized (Nespor and Sandler, 1999; Dachkovsky and Sandler, 2009; Sandler, 2010). In spoken language, the vocal cords convey both words and intonation, and different intonational patterns are manifested by fluctuations in frequency of vibration of the vocal cords, sequentially conveyed. This makes it challenging to demonstrate compositionality of intonation in spoken language, though it has been claimed to exist (e.g., Hayes and Lahiri, 1991). In sign languages, intonational signals are conveyed by articulators (such as different parts of the brows and the upper and lower eyelids) that are independent of each other and of the hands, used for words. This means that compositional structure of intonation is clearly revealed by the ways these components simultaneously combine (see **Figure 5** below).

While some of the linguistic facial expressions of sign languages are similar to expressions that can also accompany speech (Scherer and Ellgring, 2007; Kidwell, 2013), there is an important difference. In sign languages, these signals are more systematic, both in form and in distribution, and there are some differences across sign languages (Zeshan, 2004; Dachkovsky et al., 2013). Our study of ISL and ASL showed that over 90% of the relevant constituents are characterized by particular linguistic facial expressions and head positions (Dachkovsky and Sandler, 2009; Dachkovsky et al., 2013; Sandler et al., accepted).⁹

The intonational system in sign languages is itself compositional. In **Figure 5A** below, we see the raised brows of a typical yes-no question, in **Figure 5B** the squint of shared information, and in **Figure 5C**, the two intonational units combined, to characterize a yes-no question about shared information, as in *Did you see that movie we talked about last week?*

The lower face is also important in sign languages, but its role is different from that of the upper face. It conveys modification of predicates, meanings such as a ‘for a long time,’ ‘carelessly,’ ‘effortlessly’ (e.g., Liddell, 1980 for ASL, Meir and Sandler, 2008 for ISL and ASL; Sutton-Spence and Woll, 1999 for British SL). It is common that such meanings are conveyed by articulations of the lower face across sign languages that have been studied for this characteristic, although the specific lower face configuration can differ across sign languages (see Meir and Sandler, 2008 for a comparison of lower face modifiers in ASL and ISL). **Figure 6** below demonstrates a mouth shape meaning ‘for a long time’ in ISL, taken from retellings by three signers of the same part of a Tweety Bird cartoon, in which the cat

and bird fall through the air from a high place (from Sandler, 2009).¹⁰

The Head

The whole head also helps to organize information structure, for example, by assuming a particular position (such as forward for questions), or by clearly changing its position to signal a prosodic boundary (Nespor and Sandler, 1999). In the latter case, the head helps to signal dependency between clauses or information units such as topic and comment (Liddell, 1980; Dachkovsky and Sandler, 2009; Sandler et al., 2011). The full sentence example in **Figure 8** below shows the head position on either side of the prosodic boundary, in this case, separating the topic from the comment in ‘The little dog that I found last week – ran away.’

The Torso

Torso displacement takes different forms, among them shift and tilt.¹¹ A shift in the direction toward which the torso is facing tracks reference and coreference in a discourse. Shift indicates a change in speaker (signer) perspective, sometimes called role shift, and is typically used for direct or indirect quotes in discourse or for what is called constructed action (Lillo-Martin, 1995, 2012; Janzen, 2004; Cormier et al., 2013). Taken together, we can say that torso shift involves assuming the perspective of a character for a stretch of discourse (Quer, 2011; Hermann and Steinbach, 2012; Lillo-Martin, 2012; Schlenker, 2017a,b). In its most overt and full form, this displacement or shift usually consists of positioning the torso so that the chest is facing in a different direction for each perspective, shown in **Figure 7**.

A tilt, in which the body faces forward but tilts at the waist to one side or the other, can indicate contrastive focus in Sign Language of the Netherlands (Crasborn and van der Kooij, 2013), and can separate constituents in a sentence, most commonly, topic and comment (Dachkovsky et al., 2013 for unrelated ISL and ASL). Torso tilt contrasting the topic from the comment in two intonational phrases is illustrated in **Figure 7**.

In general, torso movement conveys a contrast of character perspectives or of topics in the common ground. This characterization is broad, and the cross-sign language generalizations we might glean from it must still be confirmed.¹²

The Non-dominant Hand in Discourse

Like other sublexical formational elements, the non-dominant hand can be interpreted as meaning bearing, as seen in Section

⁹There has been some debate as to whether these facial signals are components of syntactic or intonational structure. My colleagues and I have argued at length that they align rhythmically with prosodic constituents, are not isomorphic with syntactic constituents, and, like spoken intonation, perform the pragmatic role of organizing information structure (e.g., Nespor and Sandler, 1999; Dachkovsky et al., 2013; Sandler, 2010; Sandler et al., accepted). As such, although they interact with syntax, like prosodic signals in any language, they are fundamentally distinct from syntax.

¹⁰The mouth has many additional roles in sign languages, such as creating iconic gestures, much as the hands do when accompanying speech (Sandler, 2009). See also Boyes-Braem and Sutton-Spence (2001) on functions of the mouth in different sign languages.

¹¹Wilbur and Patschke (1998) deal with a third kind of torso displacement in ASL: leans on the front/back axis. They show that body leans toward or away from the addressee indicate inclusion and exclusion of the addressee. The precise functions and interactions between leans, tilts, and shifts in different sign languages suggest themselves for future research.

¹²Torso shifts and tilts can be reduced to movement of the head or eye gaze only. Distribution of these reduced signals is left to future research.



FIGURE 5 | Compositionality of facial intonation in Israeli Sign Language. **(A)** Raised brows for yes/no questions and continuation/dependency. **(B)** Squint for shared but not easily accessible information. **(C)** Raised brows and squint together for a yes/no question about shared information (following Nespor and Sandler, 1999).



FIGURE 6 | Conventionalized non-intonational ISL mouth adverbial: "for a long time" (from Sandler, 2009).

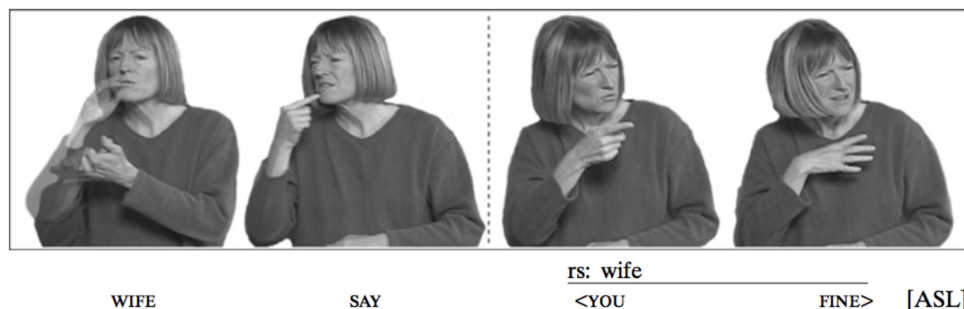


FIGURE 7 | Referential shift of upper body in American Sign Language (reprinted with permission from Lillo-Martin, 2012).

The Hands: Iconicity and Dual Duality of Patterning above.¹³ As such, it can represent a free classifier (Aronoff et al., 2003; Emmorey, 2003), or it can be dissociated from its two-handed sign, maintaining its shape and position in the signing space, and its inherent meaning, while the dominant hand goes on to produce other signs.¹⁴ An example of the latter is seen in **Figure 8**, where the non-dominant hand represents the small dog. In this way, the non-dominant hand marks different kinds of

topic continuity, disappearing from the signing space when the discourse topic changes.¹⁵

Putting the Body Back Together

If we consider the actions of the body in sign language, and work from body to linguistic structure, general properties of language stand out in high relief. The articulators each mark different linguistic functions, and they are physically independent of one another, which is also an advantage for analysis. This

¹³See Kita et al. (2014) for a discussion of the status of the non-dominant hand in signs.

¹⁴The non-dominant hand can also function as a meaningless phonological element, i.e., as a meaningless element that spreads within prosodic constituents (Nespor and Sandler, 1999).

¹⁵Space does not permit discussion of the classifier construction system here, in which the two hands can each represent a different classifier morpheme in the same expression (Supalla, 1986; Aronoff et al., 2003; Benedicto and Brentari, 2004; Sandler and Lillo-Martin, 2006; Janke and Marshall, 2017 among many others).

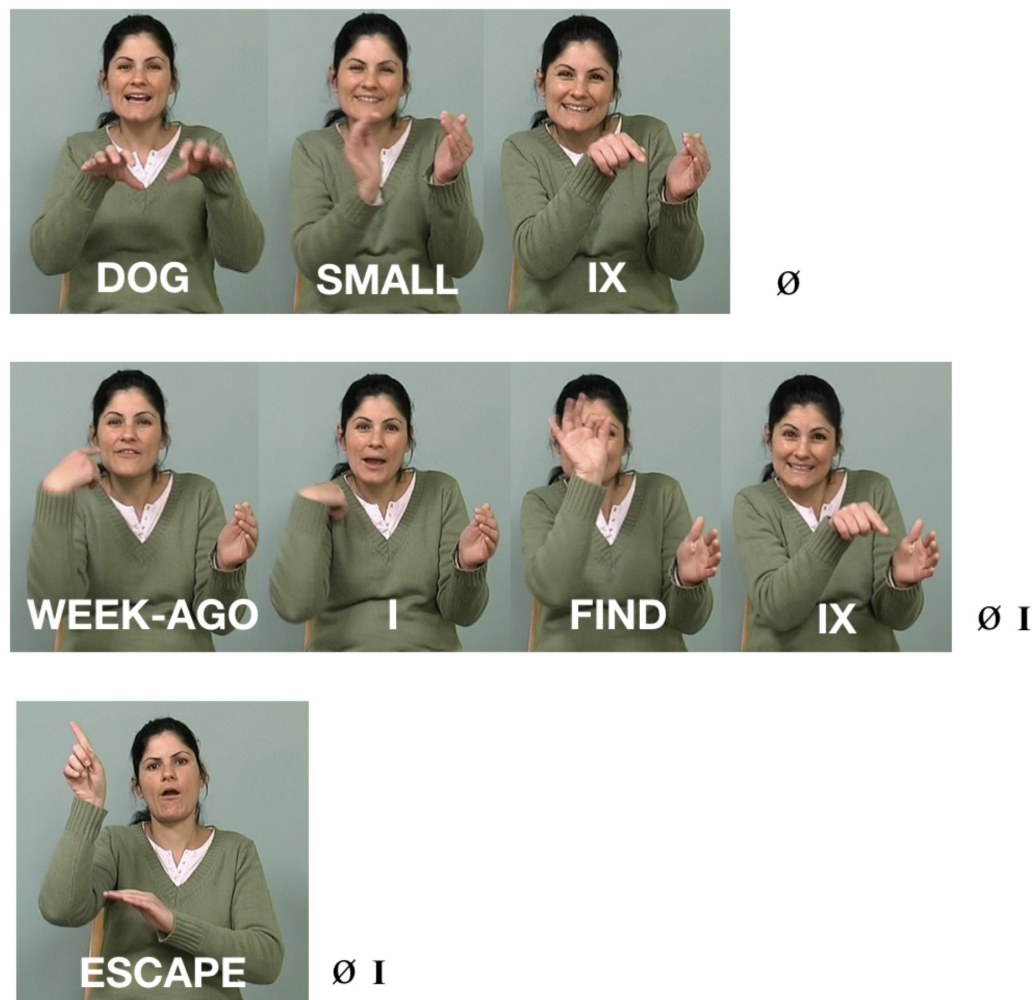


FIGURE 8 | A complex sentence in ISL, "The little dog that I found last week ran away."

independence makes it possible to incorporate a good deal of simultaneity of structure in sign language utterances, where spoken languages are much more confined to linearity. The relation between articulations and functions in sign languages is not exhaustively 1:1; the same articulation can manifest more than one linguistic function. However, when communication is exclusively visual, and is conveyed by a large number of articulators whose movements are directly perceivable and often simultaneous, the result is a system that can be both complex and transparent at the same time. This transparency often reveals general linguistic properties that are opaque or covert in spoken languages. In sign languages, complex linguistic composition can be seen at a glance.

Putting it all together, **Figure 8** above shows a sentence that means, 'The little dog that I found last week ran away.' Here is the gloss, in which 'IX' stands for 'index,' typically a pointing pronominal sign¹⁶; subscript 'I' stands for an intonational phrase;

¹⁶See Lillo-Martin and Klima (1990) and Cormier (2012) for a seminal treatment of referential loci and pronominal signs.

and subscript Ø stands for a more minor, phonological (or intermediate) phrase: [[DOG SMALL IX] Ø WEEK-AGO I FIND IX] Ø] I [[ESCAPE] Ø] I

The sentence contains two intonational phrases, separating the topic of the sentence from the comment. The first intonational phrase consists of two lower-level phonological phrases (see Nespor and Sandler, 1999). The phrases are signaled by the timing of the hands, and the facial intonation aligns itself with these prosodic constituents, as is the case with intonation patterns in spoken language.

In **Figure 8**, we see several of the characteristics described above, listed in Example (2).

(2) Composition of the complex sentence in **Figure 8**

- (a) The sign for SMALL represents dimensions and is thus two-handed;
- (b) Compositional facial expression: squint indicating shared information occurring on the entire topic ('The little dog that I found last week'); and brow raise is

added to squint at the end of the intonational phrase, signaling continuation/dependency

- (c) The head moves forward and down by the end of the topic and changes position for the comment, also marking a dependent relation between the two clauses
- (d) The topic and comment are contrasted by body tilt, which is simultaneously aligned with facial expression and head position, changing at the intonational phrase boundary.
- (e) The non-dominant hand, originating in the two-handed sign, SMALL, remains in the signing space to mark continuity of the topic ('little dog'), leaving the signing space before the comment, ('ran away'). As soon as the topic changes, the non-dominant hand configuration and location no longer signal the discourse topic (regardless of whether or not the following sign is two-handed, as it happens to be in **Figure 8**).

Culling the investigations of many sign language researchers in different sign languages over the past several decades, a Grammar of the Body in sign languages is shown in **Figure 9** below.

THE COMPOSITION OF LANGUAGE EMERGENCE

If the structure outlined above is common to sign languages generally, then we ought to be able to witness its emergence in a very young sign language. Unlike spoken languages, which are all 1000s of years old or descended from old languages, sign languages can arise at any time, and sometimes can be caught by linguists in the act of being born. Al-Sayyid Bedouin Sign Language (ABSL), a language my colleagues and I have been

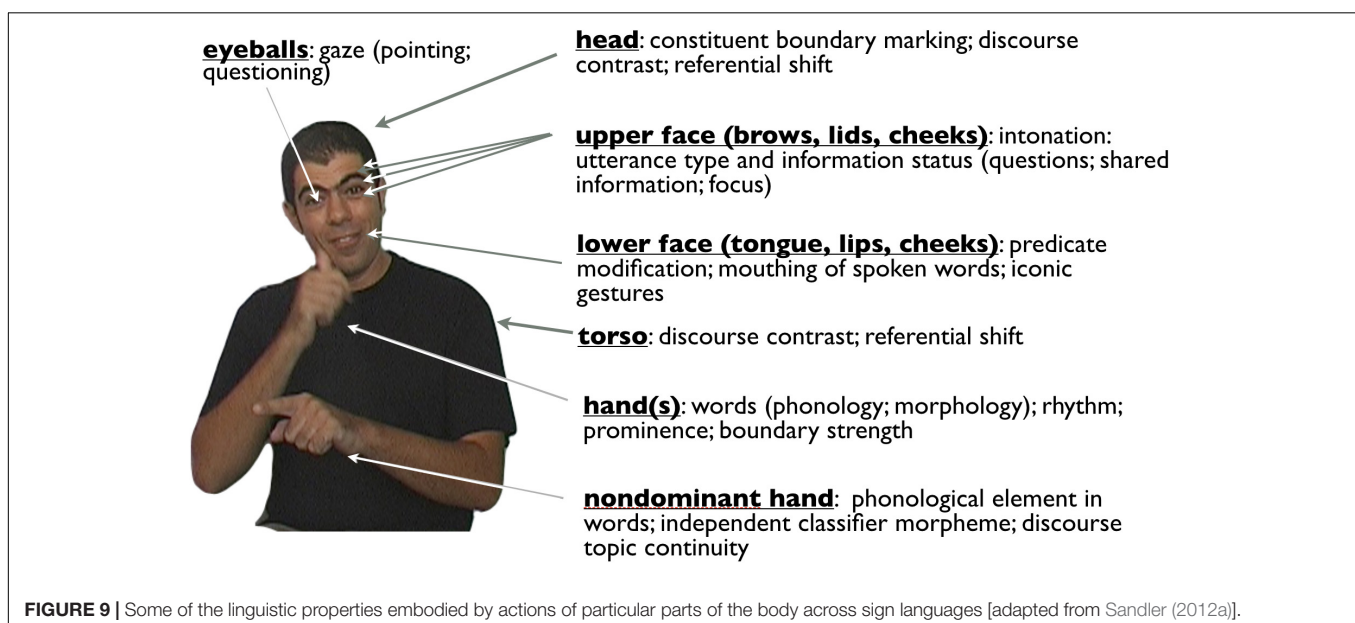
investigating, began with four deaf siblings and their family, about 90 years ago (Sandler et al., 2005, 2014). The village, today numbering about 4,000, of whom about 150 are deaf, offers the exciting possibility of uncovering the fundamental ingredients of a language and tracking their development as the language is being formed.

Al-Sayyid Bedouin Sign Language

The naïve but reasonable expectation is that new sign languages would recruit the body in a pantomimic way, so that each part of the body represents itself, 'acting out' events. A less naïve, but, as it turns out, equally wrong expectation – one that our team (Mark Aronoff, Irit Meir, Carol Padden, and myself) tacitly assumed at the outset – is this: In a community in which children have adult models and many hearing people also sign, complex linguistic structuring of the kind observed in established sign languages should arise very rapidly. Universal grammar principles and parameters ought to be hovering and beckoning, we expected, realized by children at the earliest opportunity.

We did not find this and, in particular, we did not find many linguistic structures that are widespread across established sign languages, such as a crystallized phonological system (Sandler et al., 2011), verb agreement (Padden et al., 2010), or a common type of complex classifier construction system. Instead, we found that the language began with a very simple base, but one which, crucially, bears the seeds of linguistic form, budding and blooming gradually and sporadically.

Armed with knowledge about language and sign language that our team brought with us, we were able to identify kernels of linguistic organization in syntax (Sandler et al., 2005), phonology (Sandler et al., 2011), and morphology (Padden et al., 2010; Meir et al., 2010), on their way to becoming more conventionalized and complex (see Aronoff et al., 2008; Sandler et al., 2014 for overviews). On the whole, we found that language begins with



a good deal of variation, converging on conventionalized form gradually, and at different rates for different properties (Meir and Sandler, in press).

In the analysis of two young sign languages of equal age, ABSL and ISL, I follow the outside-in paradigm that works from the body to linguistic organization, and not the traditional paradigm, which is the other way around. ABSL, a village sign language was first conceived and developed with little outside influence, unlike ISL and similar deaf community languages.¹⁷ The first generation of ABSL deaf people (four siblings) and the older members of the second generation had very little or no exposure to any other language, spoken or signed. Younger signers of the second and later generations had exposure to ISL, but the amount and quality of exposure, and of influence on their language, varies greatly, depending on the educational, family, and social environment of each individual.¹⁸ In ABSL, each age group recruits more of the body for different linguistic functions, adding complexity concomitantly in the organization of body and language.

In a videotape of a story told by one of the first four signers, already in his 60s at the time (and deceased before we began our research), the entire body is active at the outset, but not in a linguistically organized way. Only the hands are recruited linguistically, symbolizing concepts as signs. The first unit to emerge in language, then, is the word. The whole body is involved in enacting events pantomimically, so that we have a contrast in the story between HIT, a manual sign that still exists in ABSL, and 'strike,' a whole-body enactment of striking someone with a sword. With a few exceptions, each proposition in the narrative, separated by pauses, consists of a single sign representing a person, object, or action, or two-sign combinations representing a verbal expression and an argument.

In a carefully coded study of narratives of two older second-generation signers and two younger second generation signers (Sandler et al., 2011), we found that older second generation signers produce longer strings than those of the first generation man, including coordinated events as well as a rough topic-comment structure, in which constituents are separated by pause and movement of the head. Younger second generation signers add systematic, linguistic facial expressions which indicate the type of relation holding between constituents, much as the brow raise indicates dependency in the ISL example (Figure 8). In other words, recruitment of the head and then the face for non-pantomimic/affective purposes adds increasing complexity to linguistic organization as well.

The narrative of a third generation signer (Age Group IV in Tables 1, 2) is more complex still, in both body and linguistic organization. He is the son of a deaf mother and is the oldest of 5 deaf siblings. He has had considerable exposure to ISL, but can distinguish ISL from ABSL in his

TABLE 1 | Linguistic complexity in each age group [adapted from Sandler (2012a)].

Age group (Older to Younger)	Words	Complex sentences	Discourse reference/Information structure
I	Signs		
II	Signs	Unsystematic clause linking	
III	Signs	Complex sentences – embedding	Illocutionary force Parentheticals Referential shift
IV	Signs	Complex sentences – two degrees of embedding	Illocutionary force Parentheticals Double referential shift (contrasting referents) Topic continuity

TABLE 2 | Recruitment of bodily articulators for linguistic functions across age groups in ABSL [adapted from Sandler (2012a)].

Age group (Older to Younger)	Hands	Head	Face	Torso	Non-dominant hand
I	X				
II	X	X			
III	X	X	X		
IV	X	X	X	X	X

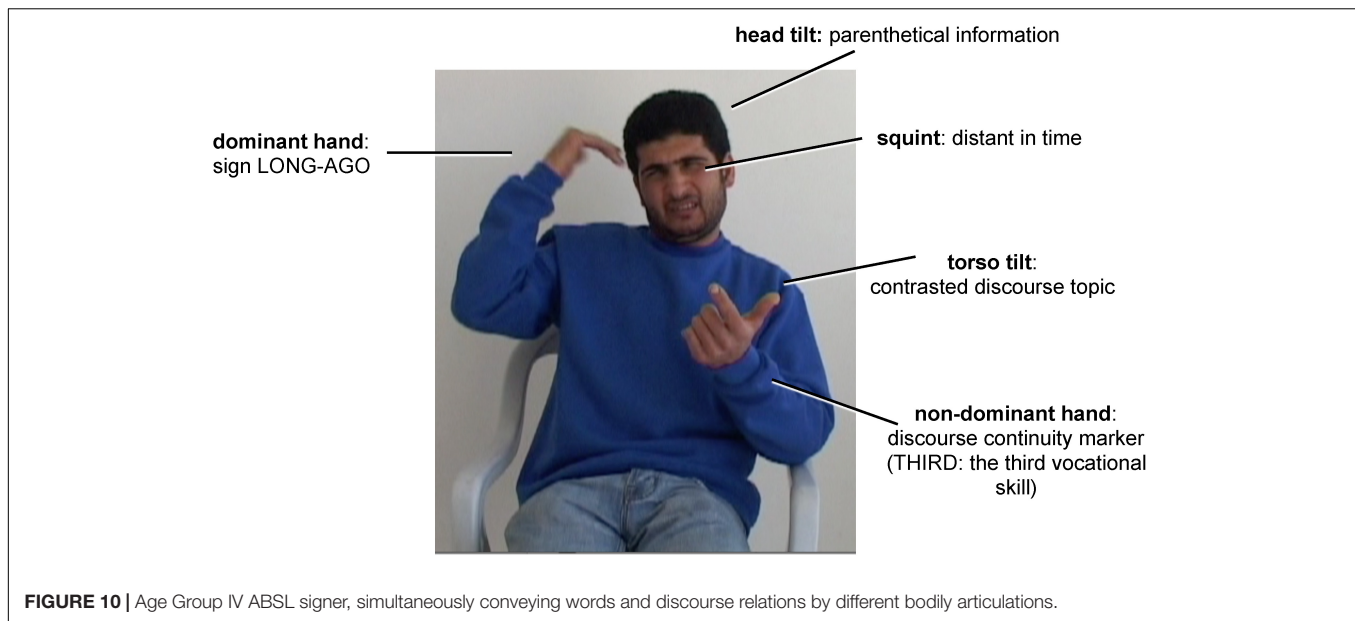
own signing, and we found only one ISL sign in his 12-min narrative. We consider him bilingual). The signer tells of his enrollment in a vocational school, and of choosing a vocation to study there. The body is divided in much the same way as in ISL Figures 8, 9. The stretch he is signing means, "The third vocation [to choose from at the vocational school] was welding. Long ago, my father was a welder. ...". The still shot in Figure 10 below was extracted from the parenthetical expression beginning with 'Long ago.' Each hand performs a different role; the head, torso, and face are independently recruited to provide relevant linguistic information in a simultaneous bodily configuration that is typical of sign languages (details in Sandler, 2012a).

Table 1 shows the overall picture of the emergence of linguistic structure in ABSL [where roman numerals refer to age groups, from oldest (I) to youngest (IV)].¹⁹ In this small but carefully documented study, there is a direct correlation between this increasing linguistic complexity and complexity in the use of the body for these functions, seen in Table 2. The hands are first – showing the human propensity for symbolization. Adding the head indicates constituency larger than the word, especially those that are connected to a following constituent, as in lists and coordinate structures. When the face is added systematically in Age Group III, illocutionary force (e.g., questions vs. declaratives) and embedding/complex sentences make their appearance. In Age Group III, the

¹⁷I adopt the distinction between village sign languages, which arise in insulated, homogeneous villages with a deaf population, and deaf community sign languages, which are conventionalized in heterogeneous populations, often in schools for deaf children (Meir et al., 2013).

¹⁸It is very difficult to find a pristine language, utterly untouched by other languages. See Meir et al. (2013) and Meir and Sandler (in press) for descriptions of the characteristics of the two language communities, and the impact of these differences on the linguistic structure.

¹⁹In all of our work on language emergence, we adopt Labov (1963) Apparent Time Hypothesis, more recently supported by Sankoff (2006). Since a person's language changes little after the critical period, we can reliably identify diachronic change by synchronically documenting the language of succeeding age groups.



torso marks larger constituents with wider scope in the discourse, distinguishing different perspectives and referents, and the non-dominant hand establishes the discourse-level topic and keeps it in the common ground. Adding articulators whose movements entail larger spatial volume contributes to more and more sophisticated structuring of a whole discourse.

We can only find the emergence of linguistic forms in new sign languages,²⁰ and we can track them most clearly by observing the recruitment of parts of the body. Were we to restrict ourselves to a model of language as computation in the mind, in which ‘externalization’ by the body is of secondary importance, we would miss these generalizations entirely.

Support From Israeli Sign Language (ISL), Another Young Sign Language

The ABSL studies rely on a small number of participants, because adult native signers of the language are so preciously few, and the results must be taken as preliminary. Israeli Sign Language is much less limited, both in the size of the deaf population (estimated at about 10,000)²¹ and in signers’ availability and flexibility. At the same time, this language arose

under very different conditions, and can be considered a Creole of many substrates but no superstrate (Meir and Sandler, 2008).²² Studies, some of them ongoing, show consistent and quantifiable correlations between the increasing organization and integration of bodily articulations and of linguistic structure in this language (Stamp and Sandler, 2016; Dachkovsky, 2017; Dachkovsky et al., accepted).

For example, Dachkovsky (2017) studied the emergence of relative clause marking across three age groups in ISL. In this language, relative clauses are marked with non-manual signals: eye squint and forward head movement. Dachkovsky found that the oldest age group often recruited only one of the markers (typically, head position) and aligned it with the noun of the relative clause alone. In a task eliciting a response corresponding to, ‘The girl who is riding a rocking horse is eating ice cream,’ older signers who produced head movement tended to align it only with the noun, ‘girl.’ The younger age group reliably recruited both markers (squint and forward head movement), and aligned them with the whole relative clause – ‘the girl who is riding a rocking horse’ – to form a constituent. The third age group performed like the second, except that the intensity of the signal was reduced, as is often the case in grammaticalization.

Another study (Dachkovsky et al., accepted) is based on spontaneous narratives, and investigates the bodily marking of discourse structures in 2 min of narrative in three age groups of ISL signers. The data were analyzed according to different degrees of discourse complexity, according to a relational hierarchy successfully used for measuring complexity and its acquisition in spoken languages. The hierarchy entails increasingly complex relations among constituents, both within and across propositions.

²⁰ A well-known example of a young sign language is Nicaraguan Sign Language (NSL, e.g., Senghas and Coppola, 2001). Some accounts of this language claim that linguistic structure emerges very rapidly (Kegl et al., 1999), and this impression is common in references to NSL, even if most NSL researchers themselves are typically more cautious. In our work on Al-Sayyid Bedouin Sign Language, we stress that its linguistic structure emerges, but gradually. It is difficult to compare NSL with either ABSL or ISL, because the methods of data elicitation and analysis are quite different.

²¹ The figure of 10,000 deaf ISL signers is considered a conservative estimation, based on enrollment in educational programs for deaf and hearing impaired children, and on figures of the health and welfare ministry regarding disability stipends (p.c. Yael Kakon, Director of the Institute for the Advancement of Deaf Persons in Israel). There are no official population figures available.

²² See Meir and Sandler (in press) for a comparison of variation and conventionalization in these two languages.

By comparing the bodily coding with discourse relations expressed, the study found that younger signers convey significantly more complex relations than older signers, and that the organization of the body to mark relations becomes more systematic. For a given relation, older signers are more likely to use different bodily markers, alone or configured together with another marker – specifically, tilt or shift of head or torso, alone or in combination – with no consistency. For the same relation, younger signers show a striking tendency to converge on a single articulator (either the head or the torso) and position, potentially freeing up other articulators to mark a different relation simultaneously. Reduction effects are also discernible in younger signers, and their distribution is still being analyzed.

By studying language emergence with the body as evidence, we have been able to arrive at generalizations regarding increases both in systematicity and in explicit marking of distinctions among different linguistic functions and relations, as a language develops over time. The generalizations described here, regarding higher levels of structure, are the only empirical evidence of language emergence available, and the Grammar of the Body provides a point of entry into the process.

What Is Compositional About It?

What is compositional about this Grammar of the Body? The compositionality principle is restated for convenience:

(1') The compositionality principle (Szabó, 2012, p. 71).

The meaning of a complex expression is determined by the meanings its constituents have individually and the way those constituents are combined.

What we have not yet demonstrated clearly is that the components are recombinable, adding predictable meaning in each recombination (see Talmy, 2003 on recombination). A clear example of recombination is found in the components of intonational facial expression. **Figure 5** showed that combining the raised brows of a yes–no question with the squint of shared

information renders a simultaneous manifestation of the two. Similarly, adding shared information to the lowered brows intonation of a wh- (content) question, renders a simultaneous combination of those two bodily expressions, as shown in **Figure 11**.

In the same way, the non-dominant hand can represent any whole sign, classifier morpheme, or numeral, and its domain is determined by the topic held in the common ground within a stretch of discourse. Torso tilts can contrast information of different kinds, from foreground/parenthetical information to different discourse referents, and the information structure determines its distribution. In these ways, the components of the body combine and recombine to convey complex information in sign languages.

It has long been observed that signers of different, unrelated sign languages can strike up a conversation and understand one another (e.g., Supalla and Webb, 1995; Newport and Supalla, 2000; Zeshan, 2015). Here we see that the use of the body, intricately orchestrated in similar ways across established sign languages, together with similar strategies for iconic symbolization, provide an envelope for understanding.

The overall picture that emerges suggests a hierarchy, in which smaller units of language are conveyed by smaller articulators and larger ones are signaled by larger (or wider reaching) articulators, schematized in **Figure 12**.

The schema is intentionally broad, in order to capture generalizations that are themselves broad. As noted, the torso conveys contrasts both between referent perspectives and between topics under discussion, and torso movement can be accompanied by or reduced to head or eye movement. The non-dominant hand, in addition to signaling topic continuity, can remain in the signing space for purely prosodic reasons, not related to meaning (Nespor and Sandler, 1999). Nevertheless, this schema captures fundamental and testable relations between language and its bodily manifestation in sign languages. Where, then, did this system come from?

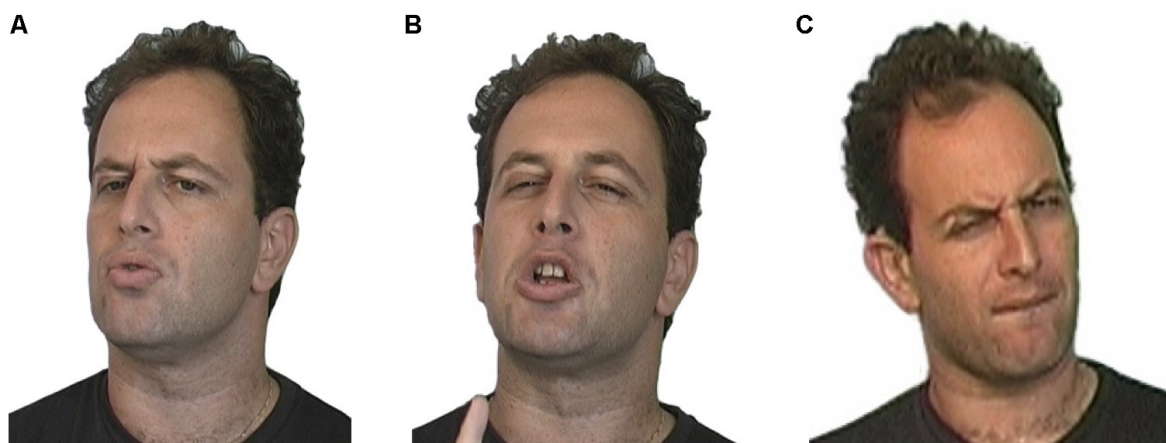
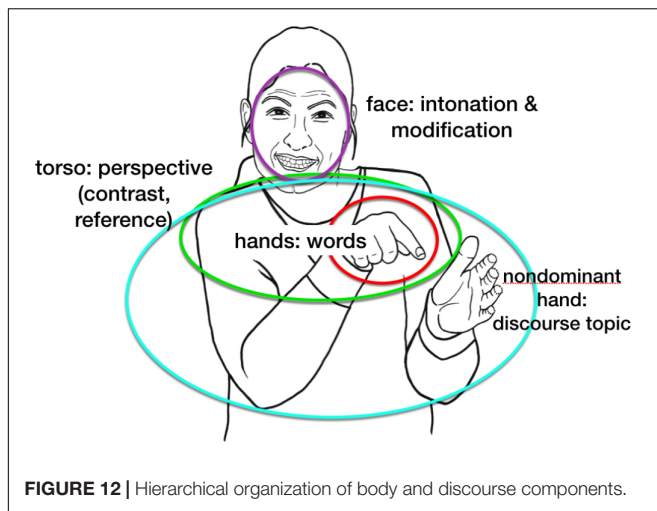


FIGURE 11 | Recombining components of facial intonation in Israeli Sign Language. **(A)** Furrowed brow for wh-question. **(B)** Squint for shared information. **(C)** Furrowed brow and squint for a wh-question about shared information (following Sandler, 1999b).



GESTURE

The flourishing field of gesture studies converges with this line of inquiry, investigating the properties both of gestures that accompany speech and of silent gesturing by hearing people in experimental tasks (e.g., Efron, 1941; McNeill, 1992; Goldin-Meadow, 2003; Kendon, 2004; Müller et al., 2013, 2014; Seyfeddinipur and Gullberg, 2014; Church et al., 2017; Goldin-Meadow and Brentari, 2017)²³. We now can state unequivocally that everyone in every culture gestures. The literature is replete with examples in which gesture adds imagistic and informative content to the message that is non-redundant – not present in the speech signal (e.g., McNeill, 1992 and many others). Most of the signal is extra-linguistic in the strict sense that it typically does not explicitly interact with the grammar, although it can be influenced by specific languages (Kita and Özyürek, 2003; Gullberg, 2011). That such elaboration is part of our universal language faculty is confirmed by the fact that signers, whose hands are occupied in the transmission of words, simultaneously produce iconic gestures with their mouths (Sandler, 2009).

Gestures can also interact more directly with linguistic structure in some cases. For example, pointing gestures are intimately integrated into speech, so that the referent of a deictic expression such as *that chair*, is unclear without pointing (Fricke, 2014).²⁴ Similarly, speakers, like signers, can set up topics in space and refer back to them with gesture, or use body position for reference. In a treatment of speech act control, Landau (2016) suggests that reference can be specified solely by the position of the body. A scenario he proposes is shown in (3). In the example, the body position (body shift, when the speaker turns to the girls) identifies the intended addressees.

(3) Body shift to evaluate addressee (Landau, 2016)

²³See Müller, 2018 for a current detailed comparison of sign and gesture.

²⁴Speakers can point with many parts of the body, not just a hand, but in true deixis, they must use a visible gesture of some sort.

Dad and mom are reading in the living room. Jen, the older daughter, is there too. The little boys and the little girls are in the kids' room, making a hell of a lot of noise. Dad tells Jen to go tell the *boys* to be quiet. Mom tells Jen to go tell the *girls* to be quiet (they are not aware of each other's orders). Jen walks over to the kids' room and says: "[To the boys:] Dad said to be quiet, [turning to the girls] and mom did too."

Landau's semantic analysis attributes the addressee function to linguistic structure, but assumes that evaluation of the addressees (the boys and the girls, respectively) belongs to extralinguistic pragmatic context. However, his analysis overlooks the fact that the body gesture is a visible signal and as such it is part of the utterance (see Kendon, 2004). In that sense, it differs from ambient pragmatic knowledge that is not signaled. Therefore, there is another possible interpretation: that bodily gesture is part of the linguistic expression. This shift in body position is reminiscent of role shift in sign languages (see the section on torso, above).

Most gesture studies attend exclusively to the hands. One exception is Birdwhistell (1970), who suggests that actions of the head and torso indicate person (first or second), prosodic boundaries, and other functions (see Kidwell, 2013 for an overview). Another is Calbris' detailed study of Parisian gestures (Calbris, 1990), which shows that the face and the hands can separately contribute information to a configuration containing both. This is the essence of compositionality.

Adam Kendon, the father of contemporary gesture studies, in early work, clearly proposes that the physical domain of gesture is the entire body. Kendon wrote:

"Just as the flow of speech may be regarded as a hierarchically ordered set of units, so we may see the patterns of body motion that are associated with it as organized in a similar fashion, as if each unit of speech has its "equivalent" in body motion... Each speech-unit is distinguished by a pattern of movement and of body-part involvement in a movement. The larger the speech unit, the greater the difference in the form of movement and the body parts involved" (Kendon, 1972, pp. 204–205).

It is not hard to see a relation between use of the body in co-speech gesture and in sign languages as schematized in Figure 12. Yet clearly, the two systems are not the same. Gesture is typically optional with speech, and actions of the body rarely comprise explicitly linguistic constructions themselves, nor are they nearly as systematic and complex as they are in sign languages (McNeill, 1992; Özyürek, 2017). Moreover, gestures that accompany speech can only be fully understood with speech, which results in a complex – and, most likely, compositional – interaction between speech units and body units.

A detailed comparison between sign language and gesture would take us too far afield (but see, e.g., Janzen and Shaffer,

2002; Müller, 2018). However, the rich gestural scaffolding that Adam Kendon describes apparently taps into the same Grammar of the Body that underpins sign languages (see Kendon, 2012). Sign languages develop the components into fully fledged, rule-governed, compositional linguistic systems. Goldin-Meadow and Brentari (2017) conclude that language (in either modality) incorporates gesture, and that the two must be studied together.

ROOTS OF COMPOSITIONAL EXPRESSION IN INTENSE EMOTIONAL DISPLAYS

All established contemporary human languages, spoken or signed, have a remarkable, creative, and productive range of expression, thanks in no small part to compositionality. Is this compositional structure in human language, so faithfully manifested in the body, alone in nature? Do other species possess it? Is it part of the language faculty alone, or might it have roots in other communicative systems of our species? The next section discusses some current issues in language evolution as context. The section following that presents evidence that human expression of intense emotion has compositional characteristics, suggesting a propensity for compositional expression in humans that is far more ancient than language.

Evolution of Language: Some Key Ideas

The field of language evolution has grown to encompass a vast body of research over the past several decades. I make no attempt to do it justice here, instead offering below only a few broad comments as context.

One widely held view is that the mental computational ability of humans to produce discrete infinity, or open-endedness, in language results from recursive application of Merge, an operation that combines two syntactic units to form a new syntactic unit. Proponents of this view hold that this single property distinguishes human language (the faculty of language in the narrow sense – FLN) from communication systems of other animals (Hauser et al., 2002).

According to one view, the computational ability attributed to FLN has no evolutionary precursor and is due to a small mutation resulting in rewiring of the human brain (Chomsky, 2007). It follows that the only reasonable direction for linguistic investigation to take is to develop the best theory to characterize this ability in contemporary humans. A different paradigm accepts the centrality of FLN in language evolution, but proposes that the evolution of this mental computational ability can be traced from cognitive (not communicative) systems of other species. Seyfarth and Cheney (2014) argue that, while there is only scant evidence for hierarchical or recursive structure in communication systems of other species, there is elaborate hierarchical structure in social cognition, particularly of non-human primates, and it is this cognitive underpinning that could have provided the basis for language. In a cogent review of a recent book by Berwick and Chomsky (2016), Fitch argues that “animal cognition offers richer parallels and potential

precursors to human thought and concepts than does animal communication” (Fitch, 2017, p. 603). He reasons that if recursive computation is a cognitive capability, then it makes sense to seek its evolutionary roots in the cognitive abilities of other species.

The uniqueness of recursion as the sole property responsible for open-endedness (‘discrete infinity’) has recently been questioned by Meir (2018). She demonstrates that a different kind of open-endedness – topic open-endedness – is a defining characteristic of human language, though it is not facilitated by recursion. Topic open-endedness refers to our uniquely human ability to express an endless variety of situations, thoughts, and ideas, real or hypothetical. She argues that, while Al Sayyid Bedouin Sign Language, an emerging language, does not have syntactically marked recursion at the outset (Sandler et al., 2011), it does have all the critical properties responsible for topic open-endedness – properties that are not present in communication systems of other species – symbolization, meaning extension, predication, negation, and compositionality.

Compositionality, the property that is the focus here, is present in all languages, including very young ones like the earliest forms of ABSL (see Sandler, 2012a for sample utterances of a first generation signer). Compositionality emerges in real time in iterated learning laboratory experiments with visually perceived stimuli, in which participants tend to extract recombinable components from holistic symbol transmission, and to assign meaning to them, from “generation” to “generation” (see Smith and Kirby, 2012 for an overview).

We find robust compositionality in the bodily division of labor in sign languages and in gesture, as shown in earlier sections. In the next section, we extend the body-as-evidence approach to address the evolution of this property. Our approach contrasts conceptually with the view that the “externalization” of language by the body is of secondary importance in language evolution (Berwick and Chomsky, 2016). The body-as-evidence view is compatible with Fitch’s (2017) position that externalization is important in understanding language evolution, but for different reasons. Fitch argues that externalization by the body is important because it can provide critical clues to computation and processing required by language. Here we see the body as manifesting, and thus revealing, compositional properties of language, directly. The experiment described below explores the human propensity for compositionality in a kind of bodily expression that is far more ancient than language: intense emotion.²⁵

Corporeal Emotional Displays of Athletes and Their Interpretation

Certain emotional configurations of facial expression (Ekman, 1992) and of body posture (de Gelder et al., 2015) are reliably interpreted in the same way. This shows that they are

²⁵Some have suggested that visible bodily forms of expression – sign, gesture, or pantomime – preceded speech in evolution (Armstrong et al., 1995; Corballis, 2003; Arbib, 2012), an issue that is orthogonal to the present discussion of communicative bodily compositionality. But it should not go unnoticed in this context that in hearing as well as deaf contemporary humans, both the mouth and the hands are intricately, profusely, and simultaneously involved in communicative expression (Boyes-Braem and Sutton-Spence, 2001; Sandler, 2009).

communicative (e.g., Fridlund, 2014). But does this form of communicative expression consist of holistic gestalts of face and body displays? Or is it compositional – like language?

Some researchers hold that facial configurations in particular are holistic or non-compositional, e.g., that all the facial actions contributing to an angry face or a happy face form a conglomerate (Ekman, 1992). Others suggest that each particular action of different parts of the face contributes its own meaning, in a structure that is compositional in nature (Russell, 1997; Scherer and Ellgring, 2007). Aviezer et al. (2012, 2015) are among the few who have considered both the face and the body together. In a study of emotional displays of athletes, they found the body to be a more reliable indicator of valence (positive or negative emotion) than the face, and they concluded that the face is ambiguous (Aviezer et al., 2015).

In our own recent work, we ask a different question. Within displays of intense emotion, we ask: Is it possible to identify emotions or emotional states associated with individual face/body features that contribute to the interpretation of the overall display? To probe this question, we investigated the displays of intense emotion by athletes who have just won or lost a competition (Cavichio and Sandler, 2015; Cavichio et al., 2018).

We select such displays first because they are intense and complex, reacting to the result of a high-stakes competition in which athletes have invested a huge amount of their lives. The intensity of the displays makes coding more straightforward, and their complexity provides a rich array of features for analysis. Second, by selecting the moment at which the athletes realize that they have won or lost, we are able to study displays which are more likely to be spontaneous and genuine, and not filtered by convention.

We began by minutely coding facial and bodily features of displays in over 300 photographs, using the Facial Action Coding System (FACS, Ekman and Friesen, 1978) for face, and a body coding system that we created. In our first study (Cavichio and Sandler, 2015), we identified the features which statistically cluster together in victory displays and in defeat displays, respectively, revealing displays prototypical of each. Our second study (Cavichio et al., 2018) presented participants with a total of 184 photographs of athletes: 49 displaying prototypical victory displays, 58 with prototypical defeat displays, 36 with ‘mixed’ displays, and 41 photos of athletes in non-competitive contexts, displaying neutral face and body.

We asked 84 participants to identify emotions or emotional states and their intensities in each display. Specifically, they were asked on a sliding scale of 0 to 100: “To what extent does the person in the image feel submissive/ashamed/sad/disappointed/frustrated/angry/happy/proud/dominant?”

We found that the most salient categories were dominance and submission, each associated with its own block of face and body features which complemented each other in the two major categories. Dominance judgments correlated with upright posture, contracted upper face, mouth open and stretched, and clenched fists (see **Figure 13**). Submission correlated with prostrate posture (kneeling or lying down), head down, face covered by the hands or otherwise not visible.

Within these broad conglomerates, positive or negative emotions could be identified by looking at individual features or feature groupings. For example, [lip corners up] (smiling mouth) was deemed happy or proud and [lip corners down] was associated with the negative emotions: sadness, disappointment, frustration, and anger. We found that [lip corners down] distinguished those negative emotions from other negative emotions which express resignation and did not have this feature: submissiveness and shame.

Individual features related to the position of the upper body also contribute to interpretation. The feature [forward upper body] was associated with the emotions submission, shame, and sadness. These emotions are grouped by Ortony et al. (1990) as evaluative disapproval and focusing on self, which we interpret as resignation. The feature [asymmetrical upper body] was significantly associated with emotions related to disappointment, frustration, and anger, grouped by the same authors as reactions to goal obstruction. The position of the upper body, then, distinguishes resignation from resistance to goal obstruction.

Our results were tallied statistically from complex displays, and included only emotions rated by participants as strongly expressing a given emotion on the emotion scale.²⁶ While the pictures of athletes were complex and did not necessarily reflect all typical constellations together, we can infer that the strongest dominant postures were typically characterized by all highly rated features together, and this is confirmed by the earlier analysis of these pictures in terms of features that clustered with victory (a typically dominant display) and similarly with defeat (a typically submissive display). Based on the findings in Cavichio et al. (2018), we have now created computer-generated 3D images that reflect abstract representations of emotional states consisting of all the features that were significantly associated with them, and that pinpoint distinctions and refinements made by individual features or feature groupings on this basis.

Figure 13 below shows images of displays that are (A) dominant, (B) dominant and happy, (C) dominant and angry, and (D) submissive and resigned. The main features associated with each one, and distinguishing them from each other and other displays as elaborated in Cavichio et al. (2018), are listed in the figure caption. We can think of these images, derived from participant ratings, as composite realizations of typical mental representations of these emotional states.

Compositionality of emotional displays reveals ancient underpinnings of compositional communication that are potentially relevant to the evolution of language. However, the use of the body in emotional displays does not correspond in any direct way to its use in language. Our results do not suggest that the Grammar of the Body sketched in relation to sign languages corresponds to the use of the body in the expression of emotion, nor would we expect it to. Language is not emotion. What the two have in common is communicativeness and complex compositionality not found to date in other species.

²⁶In our analysis, we collapsed FACS action units which overwhelmingly occurred together, such as those that lower the brows and narrow the eye aperture, referred to here as ‘contracted upper face.’ See details in Cavichio et al. (2018).

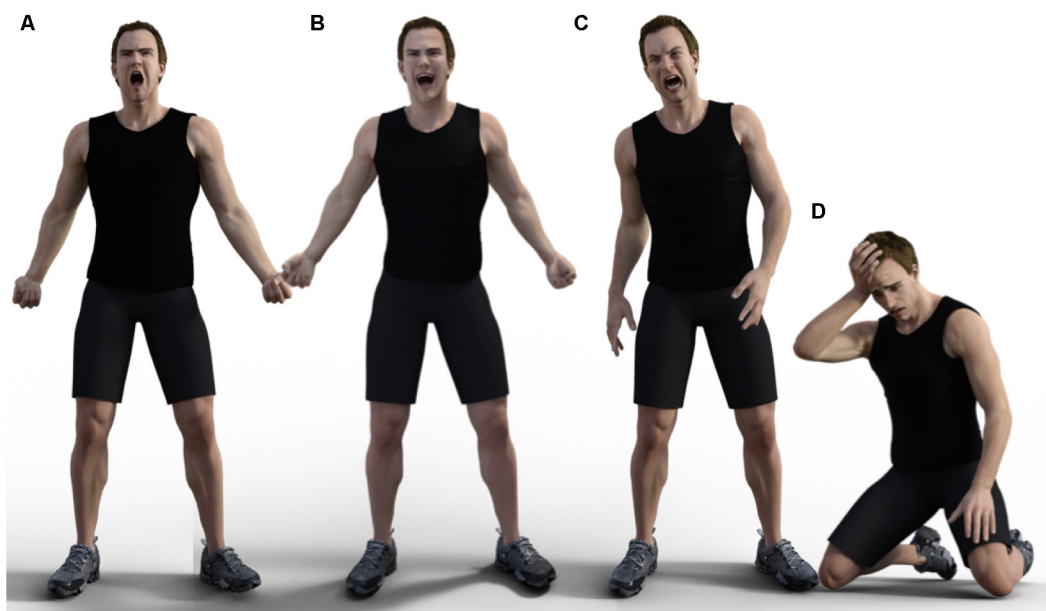


FIGURE 13 | (A) dominant: [standing, upper body back, stretched mouth, contracted upper face]. (B) Dominant plus happy, same as (A) plus [lip corners up] and [shoulders back]. (C) Dominant plus angry, same as (A) plus [lip corners down] and [asymmetric upper body]. (D) submissive: [kneeling, upper body forward, hand/s to head or face, mouth slightly open, lip corners down, inner brows raised]. Figures created by Daniel Landau.

Interpretation of emotional displays is highly context dependent. In an experiment in which actors performed contextualized narratives with nonce speech, Dael et al. (2012) found that body forward signals hot anger, while in our studies of sports competitions, torso forward is associated with resignation and submission. The difference might be attributable to differences in coding categories (whether or not ‘forward’ entails bending at the waist), or to different interpretations of the same feature in different contexts. The answer awaits future research. The interpretation of linguistic expressions is also somewhat dependent on context, but conceivably to a much lesser extent.

The complex emotional expressions described above bear the human trait of compositionality, and differ strikingly from communicative expressions of other species, as far as we know (see footnote 4). We do not yet know whether there are constraints on the combinations of face and body actions, nor do we know how productively the components that comprise them can be manipulated and recombined to form new messages. Such possibilities and comparisons of different kinds of compositional communication offer a new spectrum of research possibilities.

SUMMARY AND CONCLUSION

The relation between mind and body has been debated by philosophers for centuries (Robinson, 2017), because the issue is central to understanding human nature. Scientific investigations of spoken and signed language include the description of bodily articulation (particularly in phonology in spoken language, and of

the whole body in sign language). But since the language faculty is often seen as a property of mind alone, the role of the body is viewed as secondary for understanding the essential principles governing language. Here we propose a change, by showing that the body does provide evidence for key properties of language and its emergence.

If successful, the approach proposed here will encourage several directions of research, some of them already underway. A nuanced theory of the Grammar of the Body will make informed predictions, which can be empirically tested, about structures that are likely to occur in all established sign languages, and will uncover differences as well. Such structures can also reflect the underlying composition of spoken constituents, as we have seen in connection with sign language and gesture – from the semantic components of words to reference, complex propositions, and higher levels of discourse. Detailed comparisons between sign languages and their gestural roots, to some extent shared by all, can ensue, following Kendon’s insights (see Gesture).

Sign languages provide contemporary, empirical evidence for language emergence, in populations of contemporary humans. These emerging sign languages are the only empirical source of evidence for identifying the bare essentials of language that emerge without any model and for the development and conventionalization of complex structures across generations. The Grammar of the Body model, and more refined measures of body and language efficiency and complexity sketched in section Support From Israeli Sign Language (ISL), Another Young Sign Language, can be developed and elaborated to explore the emergence of other sign languages and their development over time.

There is no doubt that visible bodily actions evolved as part of our communicative endowment, and evolutionary biologists take the body seriously in understanding language evolution (e.g., Donald, 1993; Fitch, 2010, 2017). We are now developing a test of our preliminary findings about the compositionality of bodily displays of emotion by experimentally manipulating the components and investigating the resulting interpretations. The role of context in organizing and interpreting emotion displays vs. linguistic expressions also offers fertile ground for future comparison and characterization of these systems.

Taken together, evidence from spoken language, sign language, language emergence, co-speech gesture, and the communicative expression of emotion demonstrates that compositional communication in all domains is an inherent human trait. We have been able to arrive at this conclusion by admitting the body as evidence for the nature of language.

ETHICS STATEMENT

The research was fully approved by the European Research Council's Ethics Committee.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

- Arbib, M. A. (2012). *How the Brain Got Language: The Mirror System Hypothesis*, Vol. 16, Oxford: Oxford University Press. doi: 10.1093/acprof:osobl/9780199896684.001.0001
- Archangeli, D., and Pulleyblank, D. (1996). *Grounded Phonology*. Cambridge, MA: MIT Press.
- Armstrong, D. F., Stokoe, W. C., and Wilcox, S. E. (1995). *Gesture and the Nature of Language*. Cambridge: Cambridge University Press doi: 10.1017/CBO9780511620911
- Arnold, K., and Zuberbühler, K. (2012). Call combinations in monkeys: compositional or idiomatic expressions? *Brain Lang.* 120, 303–309. doi: 10.1016/j.bandl.2011.10.001
- Aronoff, M. (1992). "Segmentalism in linguistics: the alphabetic basis of phonological theory," in *The Linguistics of Literacy*, Vol. 21, eds P. A. Downing, S. D. Lima, and M. Noonan (Amsterdam: John Benjamins Publishing), 71–81.
- Aronoff, M., Meir, I., Padden, C., and Sandler, W. (2003). *Classifier Complexes and Morphology in Two Sign Languages*, ed. K. Emmorey (Mahwah, NJ: Lawrence Erlbaum Associates), 53–84.
- Aronoff, M., Meir, I., Padden, C. A., and Sandler, W. (2008). The roots of linguistic organization in a new language. *Spec. Issue Interact. Stud.* 9, 133–153
- Aronoff, M., Meir, I., and Sandler, W. (2005). The paradox of sign language morphology. *Language* 81, 301–344. doi: 10.1353/lan.2005.0043
- Aviezer, H., Messinger, D. S., Zangvil, S., Mattson, W. I., Gangi, D. N., and Todorov, A. (2015). Thrill of victory or agony of defeat? perceivers fail to utilize information in facial movements. *Emotion* 15, 791–797. doi: 10.1037/emo0000073
- Aviezer, H., Trope, Y., and Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science* 338, 1225–1229. doi: 10.1126/science.1224313
- Battison, R. (1978). *Lexical Borrowing in American Sign Language*. Silver Spring: Linstok Press.
- Benedicto, E., and Brentari, D. (2004). Where did all the arguments go? argument-changing properties of classifiers in ASL. *Nat. Lang. Linguist. Theory* 22, 743–810. doi: 10.1007/s11049-003-4698-2
- Berwick, R. C., and Chomsky, N. (2016). *Why only us: Language and Evolution*. Cambridge MA: MIT Press.
- Birdwhistell, R. L. (1970). *Kinesics and Context: Essays on Body Motion Communication*. Philadelphia, PA: University of Pennsylvania Press.
- Blasi, D. E., Wichman, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proc. Natl. Acad. Sci. U.S.A.* 113, 10818–10823. doi: 10.1073/pnas.1605782113
- Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486357
- Blevins, J. (2012). Duality of patterning: absolute universal or statistical tendency? *Lang. Cogn.* 4, 275–296. doi: 10.1515/langcog-2012-0016
- Bloomfield, L. (1933). *Language*. New York, NY: Holt, Rinehart and Winston.
- Brentari, D. (1998). *A Prosodic Model of Sign Language Phonology*. Cambridge, MA: MIT Press.
- Browman, C. P., and Goldstein, L. (1992). Articulatory phonology: an overview. *Phonetica* 49, 155–180. doi: 10.1159/000261913
- Calbris, G. (1990). *The Semiotics of French Gestures*. Bloomington, IN: Indiana University Press.
- Cavichio, F., Dachkovsky, S., Limor, L., Shamay-Tsoury, S., and Sandler, W. (2018). Compositionality in the language of emotion. *PLoS One* 13:e0201970. doi: 10.1371/journal.pone.0201970
- Cavichio, F., and Sandler, W. (2015). "Towards a theory of compositionality in displays of extreme emotion," *Proceedings of Gesture and Speech in Interaction (GESPIN 4)* (Nantes: University of Nantes).
- Chomsky, N. (2007). Of minds and language. *Biolinguistics* 1, 9–27.
- Church, R. B., Alibali, M. W., and Kelly, S. D. (2017). *Why Gesture? How the Hands Function in Speaking, Thinking and Communicating*. Philadelphia, PA: John Benjamins. doi: 10.1075/gs.7

FUNDING

This research has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme, grant agreement No. 340140. Principal Investigator: WS. Research reviewed in this article was funded by grants from the U.S. National Institute of Health and the Israel Science Foundation.

ACKNOWLEDGMENTS

I am especially grateful to Mark Aronoff for incisive and constructive comments, conceptual and editorial. Thanks also for useful feedback from participants at the Third GRAMBY Workshop, University of Haifa, March, 2017. I thank coinvestigators cited throughout the article and all members of the Sign Language Research Lab at the University of Haifa, especially to my close colleague and friend, Irit Meir, who passed away in February of this year, and extend my gratitude to the ISL and ABSL deaf communities. Reviewer comments prompted additions and clarifications in the article, which improved it. Thank you to Shai Davidi for video and other technical assistance, and to Shiri Barnhart for her administrative help. Illustrations were created by Debi Menashe, and 3D athlete images were created by Daniel Landau and his team.

- Coerts, J. (1992). *Nonmanual Grammatical Markers: an Analysis of Interrogatives, Negations, and Topicalisations in Sign Language of the Netherlands*. Doctoral dissertation, University of Amsterdam: Amsterdam.
- Corballis, M. C. (2003). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.
- Cormier, K. (2012). "Pronouns," in *Sign Language: An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll. Berlin: Mouton de Gruyter, 227–244.
- Cormier, K., Smith, S., and Zwets, M. (2013). Framing constructed action in British sign language narratives. *J. Pragmat.* 55, 119–139. doi: 10.1016/j.pragma.2013.06.002
- Crasborn, O. (2011). "The other hand in sign language phonology," in *The Blackwell Companion to Phonology*, Vol. 1, ed. M. Van Oostendorp, C. J. Ewen, E. V. Hume, and K. Rice (Hoboken, NJ: John Wiley & Sons), 223–240.
- Crasborn, O., and van der Kooij, E. (2013). The phonology of focus in sign language of the Netherlands. *J. Linguist.* 49, 515–565. doi: 10.1044/1092-4388(2010/08-0033)
- Dachkovsky, S. (2017). *Grammaticalization of Intonation in Israeli Sign Language*. Doctoral dissertation, University of Haifa: Haifa, IL.
- Dachkovsky, S., Healy, C., and Sandler, W. (2013). Visual intonation in two sign languages. *Phonology* 30, 211–252. doi: 10.1017/S0952675713000122
- Dachkovsky, S., and Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Lang. Speech* 52, 287–314. doi: 10.1177/0023830909103175
- Dachkovsky, S., Stamp, R., and Sandler, W. (accepted). Constructing complexity in a young sign language. *Front. Psychol.* 9:2202. doi: 10.3389/fpsyg.2018.02202
- Dael, N., Mortillaro, M., and Scherer, K. R. (2012). Emotion expression in body action and posture. *Emotion* 12, 1085–1101. doi: 10.1037/a0025737
- de Gelder, B., de Borst, A. W., and Watson, R. (2015). The perception of emotion in body expressions. *Wiley Interdiscip. Rev. Cogn. Sci.* 6, 149–158. doi: 10.1002/wcs.1335
- De Saussure, F. (1959). *Course in General Linguistics*, eds C. Bally, A. Sechehaye, A. Riedlinger, and W. Baskin. New York, NY: McGraw Hill Book Company.
- Dingemanse, M., Blasi, D. E., Lupyán, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends Cogn. Sci.* 19, 603–615. doi: 10.1016/j.tics.2015.07.013
- Donald, M. (1993). Précis of Origins of the modern mind: three stages in the evolution of culture and cognition. *Behav. Brain Sci.* 16, 737–748. doi: 10.1017/S0140525X00032647
- Downing, P. A., Lima, S. D., and Noonan, M. (1992). *The Linguistics of Literacy*, Vol. 21, Philadelphia, PA: John Benjamins Publishing. doi: 10.1075/tsl.21
- Efron, D. (1941). *Gesture and Environment*. New York, NY: King's Crown Press.
- Ekman, P. (1992). Are there basic emotions? *Psychol. Rev.* 99, 550–553.
- Ekman, P., and Friesen, W. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Palo Alto, CA: Consulting Psychologists Press.
- Emmorey, K. (ed.). (2003). *Perspectives in Classifier Constructions*. Mahwah, NJ: Lawrence Erlbaum Associates. doi: 10.4324/9781410607447
- Enfield, N. J. (2004). On linear segmentation and combinatorics in co-speech gesture: a symmetry-dominance construction in Lao fish trap descriptions. *Semiotica* 149, 57–123. doi: 10.1515/semi.2004.038
- Engberg-Pedersen, E. (1993). *Space in Danish Sign Language: The Semantics and Morphosyntax of the Use of Space in a Visual Language*. Hamburg: Signum-Verlag.
- Fernald, T. B., and Napoli, D. J. (2000). Exploitation of morphological possibilities in signed languages: comparison of American Sign Language with English. *Sign Lang. Linguist.* 3, 3–58. doi: 10.1075/sll.3.1.03fer
- Fitch, W. T. (2010). *The Evolution of Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511817779
- Fitch, W. T. (2017). On externalization and cognitive continuity in language evolution. *Mind Lang.* 32, 597–606. doi: 10.1111/mila.12162
- Frege, G. (1914/1979). "Logic in mathematics," in *Gottlob Frege: Posthumous Writings*, eds H. Hermes, F. Kambartel, and F. Kaulbach, Trans. P. Long and R. White (Chicago, IL: University of Chicago Press), 72–84.
- Fricke, E. (2014). "Deixis, gesture, and embodiment from a linguistic point of view," in *Body–Language–Communication. An International Handbook on Multimodality in Human Interaction*, 2nd Edn, eds C. Müller, S. Ladewig, A. Cienki, E. Fricke, J. Bressemer and D. McNeill (Berlin: De Gruyter Mouton), 1802–1823.
- Fridlund, A. J. (2014). *Human Facial Expression: An Evolutionary View*. San Diego, CA: Academic Press.
- Frishberg, N. (1975). Arbitrariness and iconicity: historical change in American Sign Language. *Language* 51, 696–719. doi: 10.2307/412894
- Goldin-Meadow, S. (2003). *Hearing Gesture: How our Hands Help us Think*. Cambridge, MA: Harvard University Press.
- Goldin-Meadow, S., and Brentari, D. (2017). Gesture, sign and language: coming of age of sign language and gesture studies. *Behav. Brain Sci.* 40, 1–82. doi: 10.1017/S0140525X15001247
- Gough, P. B., and Hillinger, M. L. (1980). Learning to read: an unnatural act. *Bull. Orton Soc.* 30, 179–196. doi: 10.1007/BF02653717
- Gullberg, M. (2011). "Multilingual multimodality: communicative difficulties and their solutions in second-language use," in *Embodied Interaction: Language and Body in the Material World*, eds J. Streeck, C. Goodwin, and C. LeBaron (Cambridge: Cambridge University Press), 137–151.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511616983
- Hauser, M., Chomsky, N., and Fitch, T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569
- Hayes, B., and Lahiri, A. (1991). Bengali intonational phonology. *Nat. Lang. Linguist. Theory* 9, 47–96. doi: 10.1007/BF00133326
- Hermann, A., and Steinbach, M. (2012). "Quotation in sign languages – a visible context shift," in *Quotatives: Cross-linguistic and Cross Disciplinary Perspectives*, eds I. Buchstaller, and I. Van Alphen (Amsterdam: Benjamins), 203–228. doi: 10.1075/celcr.15.12her
- Hinzen, W., Machery, E., and Werning, M. (eds). (2012). *The Oxford Handbook of Compositionality*. Oxford: Oxford University Press.
- Hockett, C. F. (1960). The origin of speech. *Sci. Am.* 203, 88–97. doi: 10.1038/scientificamerican0960-88
- Iverson, J. M., and Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature* 396:228. doi: 10.1038/24300
- Jackendoff, R. (1990). *Semantic Structures*, Vol. 18, Cambridge MA: MIT Press.
- Janke, V., and Marshall, C. R. (2017). Using the hands to represent objects in space: gesture as a substrate for signed language acquisition. *Front. Psychol.* 8:2007. doi: 10.3389/fpsyg.2017.02007
- Janzen, T. (2004). Space rotation, perspective shift, and verb morphology in ASL. *Cogn. Linguist.* 15, 149–174. doi: 10.1515/cogl.2004.006
- Janzen, T., and Shaffer, B. (2002). "Gesture as the substrate in the process of ASL grammaticization," in *Modality and Structure in Signed and Spoken Languages*, eds R. P. Meier, K. Cormier, and D. Quinto-Pozos (Cambridge: Cambridge University Press), 199–223.
- Johnston, T., and Schembri, A. C. (1999). On defining lexeme in a signed language. *Sign Lang. Linguist.* 2, 115–185. doi: 10.1075/sll.2.2.03joh
- Kegl, J., Senghas, A., and Coppola, M. (1999). "Creation through contact: Sign language emergence and sign language change in Nicaragua," in *Language Creation and Language Change: Creolization, Diachrony, and Development*, ed. M. DeGraff (Cambridge, MA: MIT Press), 179–237.
- Kendon, A. (1972). "Some relationships between body motion and speech," in *Studies in Dyadic Communication*, eds A. W. Siegman and B. Pope (New York, NY: Elsevier), 177–210.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572
- Kendon, A. (2012). Reflections on "Dedicated gestures and the emergence of sign language" by Wendy Sandler. *Gesture* 12, 308–326. doi: 10.1075/gest.12.3.02ken
- Kidwell, M. (2013). "Framing, grounding, and coordinating conversational interaction: posture, gaze, facial expression, and movement in space," in *Body–Language–Communication. An International Handbook on Multimodality in Human Interaction*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and S. Teßendorf (Berlin: De Gruyter Mouton), 100–112.
- Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* 48, 16–32. doi: 10.1016/S0749-596X(02)00505-3

- Kita, S., van Gijn, I., and van der Hulst, H. (2014). The non-linguistic status of the symmetry condition in signed languages: evidence from a comparison of signs and speech-accompanying representational gestures. *Sign Lang. Linguist.* 17, 215–238. doi: 10.1075/sll.17.2.04kit
- Klima, E., and Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.
- Kyle, J. G., Woll, B., Pullen, F., and Maddix, F. (1988). *Sign Language: The Study of Deaf People and Their Language*. Cambridge: Cambridge University Press.
- Labov, W. (1963). The social motivation of a sound change. *Word* 19, 273–309. doi: 10.1080/00437956.1963.11659799
- Landau, I. (2016). Speech act control: an unrecognized type. *Paper presented at the Israel Association for Theoretical Linguistics Meeting*, Jerusalem, IL.
- Lepic, R. (2015). *Motivation in Morphology: Lexical Patterns in ASL and English*. Doctoral dissertation, University of California: San Diego, CA.
- Lepic, R., Börstell, C., Belsitzman, G., and Sandler, W. (2016). Taking meaning in hand: iconic motivations in two-handed signs. *Sign Lang. Linguist.* 19, 37–81. doi: 10.1075/sll.19.1.02lep
- Liddell, S. K. (1980). *American Sign Language Syntax*, Vol. 52, The Hague: Mouton de Gruyter.
- Liddell, S. K., and Johnson, R. E. (1986). American sign language compound formation processes, lexicalization, and phonological remnants. *Nat. Lang. Linguist. Theory* 4, 445–513. doi: 10.1007/BF00134470
- Liebal, K., Waller, B. M., Slocombe, K. E., and Burrows, A. M. (2013). *Primate Communication: a Multimodal Approach*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139018111
- Lillo-Martin, D. (1995). “The point of view predicate in American Sign Language,” in *Language, Gesture, and Space*, eds K. Emmorey and J. S. Reilly (Hillsdale, NJ: Lawrence Erlbaum Associates), 155–170.
- Lillo-Martin, D. (2012). “Utterance reports and constructed action,” in *Sign Language: An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: Walter de Gruyter), 365–387.
- Lillo-Martin, D., and Klima, E. S. (1990). “Pointing out differences: ASL pronouns in syntactic theory,” in *Theoretical Issues in Sign Language Research*, Vol. 1, eds S. D. Fischer and P. Siple (Chicago, IL: University of Chicago Press), 191–210.
- Martinet, A. (1960). *Éléments de Linguistique Générale*. Paris: Armand Colin.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago press.
- Meir, I. (2002). A cross-modality perspective on verb agreement. *Nat. Lang. Linguist. Theory* 20, 413–450. doi: 10.1023/A:1015041113514
- Meir, I. (2010). Iconicity and metaphor: constraints on metaphoric use of iconic forms. *Language* 86, 865–896. doi: 10.1353/lan.2010.0044
- Meir, I. (2018). “Topic-Openendedness: why recursion is overrated,” in *Proceedings of the 12th International Conference on the Evolution of Language*, eds C. Cuskley, M. Flaherty, H. Little, L. McCrohon, A. Ravignani, and T. Verhoef (Edinburgh: University of Edinburgh), 295–305.
- Meir, I., Aronoff, M., Sandler, W., and Padden, C. (2010). “Sign languages and compounding,” in *Compounding*, eds S. Scalise and I. Vogel (Amsterdam: John Benjamins), 573–595.
- Meir, I., Israel, A., Sandler, W., Padden, C., and Aronoff, M. (2013). Community structure and language structure. *Linguist. Var.* 12, 247–291. doi: 10.1075/lv.12.2.04mei
- Meir, I., and Sandler, W. (2008). *A Language in Space: The Story of Israeli Sign Language*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Meir, I., and Sandler, W. (in press). “Conventionalization and variation in language emergence: the case of two young sign languages in Israel,” in *Linguistic Contact, Continuity, and Change in the Genesis of Modern Hebrew*, eds E. Doron, M. Rappaport Hovav, Y. Reshef, and M. Taube (Amsterdam: Benjamins).
- Müller, C. (2018). Gesture and sign: cataclysmic break or dynamic relations. *Front. Psychol.* 9:1651. doi: 10.3389/fpsyg.2018.01651
- Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D., Bressemer, J. (eds). (2014). *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 2, Berlin: de Gruyter.
- Müller, C., Cienki, A., Fricke, E., Ladewig, S. H., McNeill, D., and Teßendorf, S. (eds). (2013). *Body-Language–Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 1, Berlin: de Gruyter.
- Nespor, M., and Sandler, W. (1999). Prosody in Israeli sign language. *Lang. Speech* 42, 143–176. doi: 10.1177/00238309990420020201
- Newport, E. L., and Supalla, T. (2000). “Sign language research at the millennium,” in *The Signs of Language Revisited*, eds K. Emmorey and H. Lane (Mahawah, NJ: Erlbaum), 103–114.
- Ortony, A., Clore, G. L., and Collins, A. (1990). *The Cognitive Structure of Emotions*. Cambridge: Cambridge University Press.
- Özyürek, A. (2017). “Function and processing of gesture in the context of language,” in *Why Gesture?: How the Hands Function in Speaking, Thinking and Communicating*, eds R. B. Church, M. W. Alibali, and S. D. Kelly (Philadelphia, PA: John Benjamins).
- Padden, C. A., Meir, I., Aronoff, M., and Sandler, W. (2010). “The grammar of space in two new sign languages,” in *Sign Languages: A Cambridge Language Survey*, ed. D. Brentari (New York, NY: Cambridge University Press), 570–592.
- Padden, C. A., Meir, I., Hwang, S. O., Lepic, R., Seegers, S., and Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture* 13, 287–308. doi: 10.1075/gest.13.3.03pad
- Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol.* 1:227. doi: 10.3389/fpsyg.2010.00227
- Pfau, R., Steinbach, M., and Woll, B. (2012). *Sign Language: An International Handbook*. Berlin: Walter de Gruyter. doi: 10.1515/9783110261325
- Quer, J. (2011). “Reporting and quoting in signed discourse,” in *Understanding Quotation*, Vol. 7, eds E. Brendel, J. Meibauer, and M. Steinbach (Berlin: Walter de Gruyter), 277–302.
- Reilly, J., McIntire, M., and Bellugi, U. (1990). The acquisition of conditionals in American sign language: grammaticized facial expressions. *Appl. Psycholinguist.* 11, 369–392. doi: 10.1017/S0142716400009632
- Robinson, H. (2017). “Dualism,” in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Mexico: Wikimania).
- Russell, J. A. (1997). “Reading emotion from and into faces: resurrecting a dimensional-contextual perspective,” in *The Psychology of Facial Expression*, eds J. A. Russell and J. M. Fernández-Dols (Cambridge: Cambridge University Press), 295–320.
- Sandler, W. (1987). “Assimilation and feature hierarchy in ASL,” in *Chicago Linguistics Society Parasession on Autosegmental Phonology*, eds A. Bosch, B. Need, and E. Schiller (Chicago, IL: Chicago Linguistic Society), 266–278.
- Sandler, W. (1989). *Phonological Representation of the Sign: Linearity and Nonlinearity in American Sign Language*, Vol. 32, Dordrecht: Foris. doi: 10.1515/9783110250473
- Sandler, W. (1993). Hand in hand: the roles of the nondominant hand in sign language phonology. *Linguist. Rev.* 10, 337–390. doi: 10.1515/tlir.1993.10.4.337
- Sandler, W. (1999a). “Cliticization and prosodic words in a sign language,” in *Studies on the Phonological Word*, eds T. A. Hall and U. Kleinhenz (Amsterdam: Benjamins), 223–254. doi: 10.1075/cilt.174.09san
- Sandler, W. (1999b). The medium and the message: prosodic interpretation of linguistic content in sign language. *Sign Lang. Linguist.* 2, 187–216. doi: 10.1075/sll.2.2.04san
- Sandler, W. (2009). Symbiotic symbolization by hand and mouth in sign language. *Semiotica* 174, 241–275. doi: 10.1515/semi.2009.035
- Sandler, W. (2010). Prosody and syntax in sign language. *Trans. Philol. Soc.* 108, 298–328. doi: 10.1111/j.1467-968X.2010.01242.x
- Sandler, W. (2012a). Dedicated gestures in the emergence of sign language. *Gesture* 12, 265–307. doi: 10.1075/gest.12.3.01san
- Sandler, W. (2012b). The phonological organization of sign languages. *Lang. Linguist. Compass* 6, 162–182.
- Sandler, W. (2017). The challenge of sign language phonology. *Annu. Rev. Linguist.* 3, 43–63. doi: 10.1146/annurev-linguistics-011516-034122
- Sandler, W., Aronoff, M., Padden, C., and Meir, I. (2014). “Language emergence,” in *The Cambridge Handbook of Linguistic Anthropology*, eds N. Enfield, P. Kockelman, and J. Sindell (Cambridge: Cambridge University Press), 250–284. doi: 10.1017/CBO9781139342872.012
- Sandler, W., and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139163910

- Sandler, W., Lillo-Martin, D., Dachkovsky, S., and de Quadros, R. (accepted). "Prosody on the hands and face in sign language," in *The Handbook of Prosody*, eds C. Gussenhoven and A. Chen (Oxford: Oxford University Press).
- Sandler, W., Meir, I., Dachkovsky, S., Padden, C., and Aronoff, M. (2011). The emergence of complexity in prosody and syntax. *Lingua* 121, 2014–2033. doi: 10.1016/j.lingua.2011.05.007
- Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: systematic structure in a new language. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2661–2665. doi: 10.1073/pnas.0405448102
- Sankoff, G. (2006). "Age: apparent time and real time," in *Elsevier Encyclopedia of Language and Linguistics*, 2nd Edn, eds G. Sankoff and K. Brown (New York, NY: Elsevier), 110–115. doi: 10.1016/B0-08-044854-2/01479-6
- Scherer, K. R., and Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion* 7, 113–130.
- Schlenker, P. (2017a). Super-monsters I: attitude and action role shift in sign language. *Semant. Pragmat.* 10, 1–66. doi: 10.3765/sp.10.12
- Schlenker, P. (2017b). Super-monsters II: role shift, iconicity and quotation in sign language. *Semant. Pragmat.* 10, 1–67. doi: 10.3765/sp.10.12
- Schlenker, P., Chemla, E., Arnold, K., and Zuberbühler, K. (2016). Pyow-hack revisited: two analyses of putty-nosed money alarm calls. *Lingua* 171, 1–23. doi: 10.1016/j.lingua.2015.10.002
- Senghas, A. (2003). Intergenerational influence and ontogenetic development in the emergence of spatial grammar in Nicaraguan Sign Language. *Cogn. Dev.* 18, 511–531. doi: 10.1016/j.cogdev.2003.09.006
- Senghas, A., and Coppola, M. (2001). Children creating language: how Nicaraguan sign language acquired a spatial grammar. *Psychol. Sci.* 12, 323–328. doi: 10.1111/1467-9280.00359
- Seyfarth, R. M., and Cheney, D. L. (2014). The evolution of language from social cognition. *Curr. Opin. Neurobiol.* 28, 5–9. doi: 10.1016/j.conb.2014.04.003
- Seyfeddinipur, M., and Gullberg, M. (2014). *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon*. Amsterdam: John Benjamins.
- Simons, G. F., and Fenning, C. D. (2018). *Ethnologue: Languages of the World*, 21st Edn, Dallas, TX: SIL.
- Smith, K., and Kirby, S. (2012). *Compositionality and Linguistic Evolution*, eds M. Werning, W. Hinzen, and E. Machery (Oxford: Oxford University Press), 493–509.
- Stamp, R., and Sandler, W. (2016). "Articulating the emergence of language," *Proceedings of the 12th Conference Theoretical Issues in Sign Language Research (TISLR 12)* (Melbourne: La Trobe University).
- Stokoe, W. C. (1960). *Sign Language Structure: An Outline of the Visual Communication Systems of the American deaf, from the series Studies in Linguistics: Occasional Papers*. Buffalo, NY: University of Buffalo.
- Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., and Pfau, R. (2015). Event representations constrain the structure of language: sign language as a window into universally accessible linguistic biases. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5968–5973. doi: 10.1073/pnas.1423080112
- Supalla, T. (1986). "The classifier system in American Sign Language," in *Noun Classes and Categorization*, ed. C. G. Craig (Philadelphia, PA: John Benjamins), 181–214. doi: 10.1075/tsl.7.13sup
- Supalla, T., and Webb, R. (1995). "The grammar of international sign: a new look at pidgin languages," in *Language, Gesture, and Space*, eds K. Emmorey and J. S. Reilly (Hillsdale, NJ: Erlbaum), 333–352.
- Sutton-Spence, R., and Woll, B. (1999). *The Linguistics of British Sign Language: An Introduction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139167048
- Suzuki, T. N., Wheatcroft, D., and Griesser, M. (2017). Wild birds use an ordering rule to decode novel call sequences. *Curr. Biol.* 27, 2331–2336. doi: 10.1016/j.cub.2017.06.031
- Szabó, Z. (2012). *The Case for Compositionality*, eds M. Werning, W. Hinzen, and E. Machery (Oxford: Oxford University Press), 64–80.
- Talmy, L. (2003). Recombinance in the evolution of language. *Proc. Annu. Meet. Chicago Linguist. Soc.* 39, 33–66.
- Taub, S. F. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511509629
- van der Hulst, H. (1993). Units in the analysis of signs. *Phonology* 10, 209–241. doi: 10.1017/S095267570000004X
- Van der Hulst, H. (1996). On the other hand. *Lingua* 98, 121–143. doi: 10.1016/0024-3841(95)00035-6
- van der Kooij, E. (2002). *Phonological categories in Sign Language of the Netherlands*. Doctoral dissertation, Utrecht: LOT Publications.
- Wilbur, R. B. (2008). "Complex predicates involving events, time and aspect: is this why sign languages look so similar," in *Theoretical Issues in Sign Language Research*, ed. J. Quer (Hamburg: Signum Press), 217–250.
- Wilbur, R. B., and Patschke, C. G. (1998). Body leans and the marking of contrast in American Sign Language. *J. Pragmat.* 30, 275–303. doi: 10.1016/S0378-2166(98)00003-4
- Wohlgemuth, M. J., Sober, S. J., and Brainard, M. S. (2010). Linked control of syllable sequence and phonology in birdsong. *J. Neurosci.* 30, 12936–12949. doi: 10.1523/JNEUROSCI.2690-10.2010
- Wunderlich, D. (2012). "Lexical decomposition in grammar," in *The Oxford Handbook of Compositionality*, eds M. Werning, W. Hinzen, and E. Machery (Oxford: Oxford University Press), 307–327.
- Zeshan, U. (2004). Interrogative constructions in signed languages: cross-linguistic perspectives. *Language* 80, 7–39. doi: 10.1353/lan.2004.0050
- Zeshan, U. (2015). Making meaning: communication between sign languages users without a shared language. *Cogn. Linguist.* 26, 211–260. doi: 10.1515/cog-2015-0011

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer BC and handling Editor declared their shared affiliation.

Copyright © 2018 Sandler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Constructing Complexity in a Young Sign Language

Svetlana Dachkovsky, Rose Stamp* and Wendy Sandler

Sign Language Research Laboratory, University of Haifa, Haifa, Israel

OPEN ACCESS

Edited by:

Judit Gervain,
Centre National de la Recherche
Scientifique (CNRS), France

Reviewed by:

Antonio Benítez-Burraco,
Universidad de Sevilla, Spain

*Correspondence:

Rose Stamp
rose_stamp@hotmail.com

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 04 January 2018

Accepted: 24 October 2018

Published: 13 December 2018

Citation:

Dachkovsky S, Stamp R and
Sandler W (2018) Constructing
Complexity in a Young Sign
Language. *Front. Psychol.* 9:2202.
doi: 10.3389/fpsyg.2018.02202

A universally acknowledged, core property of language is its complexity, at each level of structure – sounds, words, phrases, clauses, utterances, and higher levels of discourse. How does this complexity originate and develop in a language? We cannot fully answer this question from spoken languages, since they are all thousands of years old or descended from old languages. However, sign languages of deaf communities can arise at any time and provide empirical data for testing hypotheses related to the emergence of language complexity. An added advantage of the signed modality is a correspondence between visible physical articulations and linguistic structures, providing a more transparent view of linguistic complexity and its emergence (Sandler, 2012). These essential characteristics of sign languages allow us to address the issue of emerging complexity by documenting the use of the body for linguistic purposes. We look at three types of discourse relations of increasing complexity motivated by research on spoken languages – additive, symmetric, and asymmetric (Mann and Thompson, 1988; Sanders et al., 1992). Each relation type can connect units at two different levels: within propositions (simpler) and across propositions (more complex).¹ We hypothesized that these relations provide a measure for charting the time course of emergence of complexity, from simplest to most complex, in a new sign language. We test this hypothesis on Israeli Sign Language (ISL), a young language, some of whose earliest users are still available for recording. Taking advantage of the unique relation in sign languages between bodily articulations and linguistic form, we study fifteen ISL signers from three generations, and demonstrate that the predictions indeed hold. We also find that younger signers tend to converge on more systematic marking of relations, that they use fewer articulators for a given linguistic function than older signers, and that the form of articulations becomes reduced, as the language matures. Mapping discourse relations to the bodily expression of linguistic components across age groups reveals how simpler, less constrained, and more gesture-like expressions, become language.

Keywords: language complexity, language emergence, sign languages, discourse relations, gesture, use of body, compositionality

¹Propositions are semantic constructs that usually represent meaning of sentences and discourse, and are subject to truth conditions. They consist of a predicate, a number of arguments and one or more modalities (e.g., Goddard, 2011).

INTRODUCTION

The form of language is complex at each level of structure – the word, the phrase, the clause, and higher units in the linguistic hierarchy. And at each level, language is compositional – building up complex structures by combining and recombining simpler meaningful units. Children inevitably acquire this complex system, but not all at once. The gradual, step-by-step process of acquisition offers insight into the relative complexity of different language structures and their interaction (Brown, 1973; Barrett, 2016; Dromi, 2016; Tomasello and Brooks, 2016). The contribution of the child's mind to this process is clearly impressive, but the process always occurs in the presence of adult models, which also contribute to acquisition. How does a new language accrue linguistic complexity from scratch? What are the characteristics of language emergence *de novo* in a community? Sign languages offer an opportunity to watch this phenomenon unfold.

This opportunity is unique to sign language in two ways: (1) they are young languages, and new ones can arise at any time and (2) there is often a direct correspondence between visible physical articulations and linguistic structures, providing a more transparent view of the emergence of complexity. The emergence of complexity with no model in spoken languages cannot be traced, because spoken languages are all thousands of years old or descended from old languages with complete linguistic structures. However, sign languages arise spontaneously in a community of signers, and if linguists are in the right place at the right time, they can observe the emergence of language.

It is generally assumed that sign languages begin life as gestural systems (Janzen and Shaffer, 2002; Goldin-Meadow, 2003; Meir et al., 2016) and interact with gesture even as they transform into linguistic systems (Liddell, 2003; Padden et al., 2013). Gestures are used by all of us, and they accompany spoken language. However, they are unlike language because they are less conventionalized and less systematic (McNeill, 1992). Established sign languages are clearly fully linguistic systems (Sandler and Lillo-Martin, 2006; Pfau et al., 2012). But how did they get that way?

A good deal of research has described the home sign system created by deaf children without a language model (Goldin-Meadow, 2003), and compared it with established sign language or with spoken language. Though most studies of the gestures that accompany speech are restricted to manual gestures, it has long been noted in studies of co-speech gesture that actions of the whole body accompany linguistic interaction (Kendon, 1972; Kidwell, 2013). But no study to date has attempted to combine measures of discourse complexity and bodily systematicity in order to map the emergence of language in a community, as we do here.

As will become clear, unlike the relation between form and meaning in spoken languages, many linguistic structures in sign languages have overt physical form, so that actions of different articulators – the hands, face, head, and torso – convey particular linguistic information. This is another advantage for our pursuit, as we explain. We do not claim that all aspects of sign language structure have overt physical correlates. We accept,

for example, that syntactic, semantic, and other relations and properties can be covert, and that evidence for them can be attained through linguistic analysis of the same kind that applies to spoken language (Sandler and Lillo-Martin, 2006; Pfau et al., 2012). However, the unique kind of direct mapping that we will demonstrate offers an opportunity to observe particular linguistic properties directly as they unfold.

One study to date has adopted the strategy of matching bodily form to linguistic function in a newly emerging language, Al Sayyid Bedouin Sign Language (ABSL). This language was formed in a Bedouin village in present day Israel, after four deaf children were born in a single household, and deafness began to proliferate throughout the village (Sandler et al., 2005). The study showed that with each new generation of signers, additional articulators were recruited to convey increasing linguistic complexity as the language developed (Sandler et al., 2011; Sandler, 2012). In particular, Sandler (2016) found that the first overt markings to appear served to organize discourse functions, such as topic-comment structure, referent perspective, and topic continuity across a discourse – and in that order. This approach and its preliminary findings motivate the current study, and we describe it in more detail in **Emerging Sign Languages: Use of Body Articulators**.

We investigate the emergence of complexity in a different sign language of the same age as ABSL, Israeli Sign Language (ISL). This language has developed under different social conditions from those of ABSL (described briefly in **The ISL Community and the Formation of the Language**), leading to certain differences in the path of emergence (Meir et al., 2012; Meir and Sandler, in press). In more than a decade of research on ABSL, the research team refrained from attributing complex syntactic structure to utterances without overt evidence, and instead based their analyses primarily on the meaning and prosodic structure of the productions, and their interaction (Padden et al., 2010b; Sandler et al., 2014, 2005). We follow that strategy in the present study of ISL as well.

We adopt the theory of discourse relations proposed by Mann and Thompson's (1988) and Sanders et al.'s (1992), laid out in **Measuring Relations and Complexity in Discourse**, to investigate the degree of complexity of utterances, and to measure the frequency and systematicity of the discourse relations they convey, including, for example, dependency between clauses. We look at three types of relations of increasing complexity, motivated by research on spoken languages. In the spirit of Mann and Thompson (1988) and Sanders et al. (1992, 1993), we adopt the following terms: additive, symmetric and asymmetric (see **Types and Levels of Discourse Relations and Their Relative Complexity**). Each of these types of relations can occur at two different levels: within propositions (simpler) and across propositions (more complex). We hypothesized that these constructions provide a measure for charting the emergence of complexity in a young language, from simplest to most complex, and tested the hypothesis on ISL.

By studying the recruitment of articulators to express linguistic form in fifteen ISL signers from three generations, we show in the section **Discussion: Bodily Marking Emerges Gradually** that this is indeed the case. In other words, the

emergence of constructions reflects the degree of complexity in terms of relation type and level. We also find evidence for increasing systematicity and automaticity of form as the language matures. Our conclusions and suggestions for future research comprise the **Conclusion**.

The ISL Community and the Formation of the Language

Israeli Sign Language is the established language of the deaf community in Israel (Meir and Sandler, 2008). It is a young sign language, roughly about 90 years old, which arose with the formation of the deaf community in Israel around the 1930s, beginning with the establishment of the first Israeli School for the Deaf in 1932 in Jerusalem. Immigrants from all over the world contributed to the signing used by a small number of deaf Jews and Arabs already in Jerusalem. Vocabulary items have been traced to a small number of immigrants from Germany, and immigrants from elsewhere in Europe, North Africa, and the Middle East also brought their sign languages or home sign systems with them. A conventional local sign language evolved, and today, ISL is used in a wide range of settings including the educational system, deaf social and cultural institutions, interpreting programs, and the media.

The linguistic structure of ISL is investigated in earlier work (e.g., Meir, 1998, 2010; Meir and Sandler, 2008; Meir and Sandler, in press) and its emergence has recently become the object of study, briefly noted in **The Body as a Marker of Linguistic Complexity** below.

The Body as a Marker of Linguistic Complexity

Signers exploit the use of the hands, torso, head and facial expression to convey linguistic information. Early sign language research demonstrated that non-manual signals play important roles in American Sign Language (ASL) grammar by systematically co-occurring with various linguistic structures: questions, topics, conditionals, and others (e.g., Baker and Padden, 1978; Liddell, 1978, 1980; Baker-Shenk, 1983; Reilly et al., 1990). Later, similar phenomena were demonstrated in other sign languages (Bergman, 1984, Swedish Sign Language; Engberg-Pedersen, 1990, Danish Sign Language; Coerts, 1992, Sign Language of the Netherlands (NGT); Nespor and Sandler, 1999, Israeli Sign Language; Sutton-Spence and Woll, 1999, British Sign Language; Herrmann and Steinbach, 2013, German Sign Language). Although most of this research studied facial expressions, a few studies focused on the role of other articulators, such as the head (Sandler et al., 2011; Dachkovsky et al., 2013; Lackner, 2013; Puupponen et al., 2015) and torso (Wilbur and Patschke, 1998; Ormel and Crasborn, 2012); also see Pfau and Quer (2010) and Sandler (2010) for overviews of non-manuals in sign languages.

While early work on the role of non-manual markers in various structures such as interrogatives, topics, and relative clauses attributed them to the syntactic level of analysis (e.g., Liddell, 1978, 1980; Neidle et al., 2000), other researchers have argued that facial expressions and head movements are driven

by various information structure and discourse considerations, such as topic continuity, foregrounding-backgrounding in subordinate constructions and others (Dachkovsky et al., 2013; Sandler et al., in press). For example, **Figure 1** below illustrates the marking of a neutral conditional in ISL. It typically consists of raised eyebrows and forward movement of the head (Dachkovsky and Sandler, 2009).

Moreover, those signals can co-occur with other signals to create a more complex grammatical meaning. Thus, raised brows and forward head movement signaling conditionality can combine with squinted eyes to create a more complex linguistic form – together they systematically mark counterfactual conditionals in ISL (Dachkovsky, 2008). An example of this subordinate construction using a complex array of facial expressions and head movements is presented in **Figure 2**. The antecedent, or ‘if’ clause, in this example comprises an intonational phrase, and the head position and facial intonation align with the timing of the hands to mark the phrasal boundary. Thus, the head posture (head movement forward) and facial expression (raised brows and squinted eyes) change between the last sign of the first clause and the first sign of the second clause. This change follows the manual cue at the intonational phrase boundary – the hold in position of the last sign CATCH-BALL of the first clause.

These findings motivated a study of a newly emerging Bedouin sign language, which we discuss below. The study demonstrated that, in addition to the hands, head, and face, the torso and the non-dominant hand independently can be recruited for discourse organization. Since each articulator contributes additional linguistic information, recruitment of more articulators for different functions implies more complexity of language structure.

Taking the findings across sign languages into consideration, we arrive at a very general model that relates bodily articulators to linguistic roles across sign languages (see Sandler, 2018



FIGURE 1 | Typical intonational display of the antecedent clause of an ISL neutral conditional in a sentence meaning, “If you eat now, you won’t be hungry for lunch.” The image was captured while the signer was producing the underlined sign.



FIGURE 2 | Typical intonational display of the antecedent clause of an ISL counterfactual conditional and its change in the consequence clause, in a sentence meaning, “If the goalkeeper had caught the ball, the team would have won the game” and glossed [GOALKEEPER HE CATCH-BALL] [WIN]. The figure shows the underlined signs.

for more details). This sort of correspondence is derived from a range of data and methods of analysis in different sign languages, and awaits statistical confirmation. Here our point of departure is that reliable correspondences between articulator activation and linguistic roles exist, and we test them statistically across different generations of signers in a young sign language.

Emerging Sign Languages: Use of Body Articulators

The new field of emerging sign languages has laid the foundation for understanding how language arises. Yet, in most cases, these studies evaluate particular structures at the level of the word, the clause, or the sentence, without generalizing across levels. Since sign languages are transmitted in the visual modality and use multiple articulators, the findings of most earlier studies do note the role of the body in the process of the emergence of complex grammatical distinctions, but only indirectly.

Here we aim to overcome these limitations. First, the theory of discourse relations that we adopt allows us to make broader generalizations about the emergence of complexity across the clause and the sentence levels. Second, we exploit the direct relation between bodily action and linguistic structure to evaluate the emergence of complexity and systematicity.

Senghas (1995) and Senghas et al. (1997) were pioneers in this field. They have claimed rapid language development and change between cohorts of children in a deaf school in Nicaragua. In their work, the researchers focused on the development of temporal and spatial devices in this rapidly developing language (Senghas and Coppola, 2001; Kocab et al., 2016). Assignment to a cohort reflects both the age at which the signers arrived at the newly established school, and whether or not they had signing

models in the environment. Members of the first cohort were older when they arrived at the school, and had no models for creating a language, while the second cohort were younger and had the advantage of the older cohort as a language model.

In their work, the researchers examined the emergence of particular discourse signals, often called referential shift devices – that is, devices, which shift the perspective of the discourse. In addition to lexical labels, sign languages can mark the shift with a manual point or with a movement of the body to a specified location in the three-dimensional space in front of the signer, capitalizing on the spatial affordances of the visual modality. While there was no significant difference between age cohorts of Nicaraguan Sign Language (NSL) in the use of neutral lexical signs and indexical points, there was a difference in the use of spatial devices (e.g., indexical points to space, body shifts and spatially modulated signs), with second-cohort signers using them significantly more (Kocab et al., 2015).

Ergin (2017) investigated the development of a recently discovered young sign language – Central Taurus Sign Language (CTSL), which emerged in the 1960s in a remote village situated in the mountainous region of Southern Turkey. The researcher reported on the emergence of a phonological system, handshape classifiers and argument structure in this village sign language, with a special focus on the way the semantic complexity in various different scenarios is realized on the surface structure of such a young language. In her study of the word order in this language, the author also demonstrated that the more specified use of body articulators (‘body segmentation’) in signaling reciprocal argument relations in a sentence is more characteristic of the younger signers’ production (Ergin, 2017).

The youngest reported sign language – the Sao Tome and Principe Sign Language (LGSTP) – started to emerge just a few years ago and is still in its first age cohort. The research

group investigating this language conducted a longitudinal study through a few successive sessions of video recordings. One of their findings is that the earlier stages of language development are characterized by larger signing space than subsequent, later stages (Mineiro et al., 2017), measured according to the size of the joints involved in sign production.

By and large, all the studies mentioned so far have traced the emergence of the signed word, morphological complexity, and syntax within the sentence. As noted above, their findings related to the use of the body were an artifact of studying languages conveyed by corporeal articulators. Taking linguistic functions as the point of departure, their strategy can be summarized as a function-to-body approach.

A different perspective has arisen in the studies investigating a young sign language that emerged in a Bedouin village in the Negev desert of present day Israel. ABSL has been the object of study for over a decade by Aronoff et al. (2008) and Sandler et al. (2014, for an overview). ABSL began with four deaf children in a single family about 90 years ago, and the deaf population has since spread throughout the village, now numbering about 150 deaf people in a village of 4,000. ABSL has been developing across generations of signers. In their work on this young language, the team has been especially interested in the externalization by the body of the emergence and development of grammatical functions. The researchers paid special attention to manual timing and to use of the face and head in order to understand the structuring of sentences in ABSL (Sandler et al., 2005, 2011; Padden et al., 2010a).

Taking those studies as a basis, Sandler (2012) explicitly addressed the emergence of ABSL from a body-to-function perspective. With this approach, she investigated broader discourse functions, such as the discourse topic. This preliminary study traced the step-by-step recruitment of different articulators – the face, the head, the torso, and the non-dominant hand – to create an increasingly complex linguistic system in ABSL. In this way, a correlation was found between the increase in language complexity and the affordances of multiple bodily articulators participating in language expression at higher levels of discourse.

Two observations emerge from this work, which are relevant for the present study. The first is that there is often a direct correspondence between linguistic and bodily complexity. The second is that the body traces the order of emergence of linguistic structure, such that words on the hands are first; propositions and links between them, signaled by the head and face, are next; and broader discourse organization, embodied in movement of the torso and independent spatial placement of the non-dominant hand, is last. Details of this emergence are expanded in Sandler (2012, 2013).

Although ABSL studies rely on a small number of participants, due to the exigencies of fieldwork in a community of this kind, ISL offers a field that is much less limited, both in the size of the deaf population (estimated at about 10,000) and in their availability. At the same time, ISL arose under very different conditions, and can be considered a creole of many substrates but no superstrate (Meir and Sandler, 2008), so that the stages of its emergence may be less crisply defined. Nevertheless, concrete

results about the emergence of this language have been reported. For example, conventionalization of the use of space was studied in Padden et al. (2010a). Another study found consistent and quantifiable relations between the increasing organization of bodily articulators and of linguistic structure in this language (Stamp and Sandler, 2016). Specifically, younger signers are more likely than older signers to use the head and the body simultaneously for separate linguistic functions.

In a study of relative clauses across age groups, Dachkovsky (2017) found that younger signers, unlike older signers, consistently organize this construction by aligning the noun and predicate of the relative clause with characteristic head positions and facial expressions. In the youngest group, the bodily markers are phonetically reduced, indicating the increased automaticity and conventionalization typical of grammaticalization (Dachkovsky, 2017). In general, these studies point to increased language complexity tied to increased articulatory complexity, as well as increased efficiency in use of different parts of the body as a language matures.

Here we develop the body-to-function approach and investigate the emergence of complexity and conventionalization in finer resolution. Specifically, we trace the relationship between the recruitment of bodily articulators and the complexity of discourse relations both within and across propositions. We adopt a particular measure of discourse relations and complexity, described in **Measuring Relations and Complexity in Discourse**, and investigate their emergence by tracking actions of the body for different linguistic units across age groups. We review relevant properties of language change, such as convergence as well as reduction of form in the process of conventionalization.

Measuring Relations and Complexity in Discourse

We approach the study from the perspective of relations among constituents in a discourse, and their relative complexity. By discourse, we mean a coherent multi-utterance dialog or monolog text. Discourse contains propositions, where propositions are usually understood to be truth bearing statements denoting states of affairs (e.g., Wittgenstein, 1961; Krifka, 2001; Cristofaro, 2003), but it is more than a sequence of propositions. Despite some key differences, all definitions view discourse structure as the conceptual organization of a text, driven by the communicative goals of language users, the direction of information flow and considerations of common ground.

Discourse structure subsumes such notions as segmentation, anaphoric relations, and relations between segments (Kruijff-Korbayová and Steedman, 2003). As a result, stretches of discourse are analyzed as connected to each other through a range of discourse relations (see, among others, Gernsbacher and Givon, 1995; Graesser et al., 1997; Noordman and Vonk, 1997).

In the present study we are concerned only with relational aspects of discourse organization. Discourse relations usually connect events and situations described in propositions, and,

therefore, cross the bounds of isolated propositions (cf. Hobbs, 1979; Mann and Thompson, 1986; Sanders et al., 1992, 1993). Yet, elements at a lower discourse level, within propositions, also contribute to discourse connectivity. For example, relations between topic and comment contribute substantially to information packaging.

In discourse, both explicit and implicit devices signify links between propositions, and between groups of propositions (e.g., Mann and Thompson, 1988; Kruijff-Korabayová and Steedman, 2003). Our approach encompasses both conceptual relations in discourse connectivity and overt linguistic signals. This approach is especially appropriate in the signed modality, where formal syntactic machinery for signaling these relations may be missing, controversial, and/or lacking empirical support, particularly in new sign languages.

Not all discourse relations contribute to complexity in the same way. In **Types and Levels of Discourse Relations and Their Relative Complexity and Increasing Complexity in Language Ontogeny, Diachrony, and Typology**, we introduce the types and levels of discourse organization that we adopt in terms of complexity, and briefly survey empirical evidence.

Types and Levels of Discourse Relations and Their Relative Complexity

A significant part of the discourse literature has focused on the question of how various sets of relations should be organized and what principles guide their groupings. Sanders (1997, 121) determined the properties common in all relations, in order to define “the relations among the relations” relying on the assumption that some discourse relations are more alike than others (see also Mann and Thompson, 1988). Within an organized system of discourse, its segments² may bear relation to the system as a whole, or to each other, or to both. On these grounds, some discourse relations are described as more basic and others as more complex. We adopt here the general principles of an approach that distinguishes relations based on two criteria: *types* and *levels* (Sanders et al., 1992; Givón, 1995).

The first criterion, relation *type*, distinguishes between degrees of connection between the units of discourse, ranging from additive (weakly connected) to asymmetric (strongly connected) relations. In *additive*³ relations, basic units bear relations to the system as a whole but not to each other (Sanders et al., 1992). A more complex type of relation is *symmetrical* (e.g., Cristofaro, 2008; Givón, 2009a; Nir and Berman, 2010): units of the same rank are related both to the system as a whole, and to each other. Symmetric type relations can involve either coordinate (e.g., *I walk to work and also walk back home*) or contrastive relations (e.g., *I walk to work but drive back home*). In a third type of organization, known as *asymmetric* relations, the units are not of the same rank; one unit is dependent on another unit in the system (e.g., Cristofaro, 2008; Givón, 2009a; Langacker, 2009; Nir and Berman, 2010; see **Table 1**).

The units in example (1) in **Table 1**, show additive relations within a proposition, where the only relation between them is

TABLE 1 | Examples from our data for each relation.

	Within proposition	Across proposition
Additive	(1) My father was [hammering], [sawing], [and painting].	(4) [I talked with Dani], [Kate danced with Peter], [and Sveta ate her sandwich].
Symmetric	(2) [As well as sign language (being important)], [I don't rule out the importance of spoken language].	(5) (The basketball game), it was [Tel Aviv] vs. [Haifa].
Asymmetric	(3) [My father and his wife] [they both had to stay].	(6) (Doctor said). . . [because you're not getting vitamin B12], [you need injections].

that they belong to the same whole. The order of items in *additive* relations does not change the meaning of the proposition. Example (2) is different because the units of the same rank are related *symmetrically* to each other through contrast. In (3), the elements of the utterance are not of the same rank so that the topic, ‘My father and his wife,’ serves as an anchor for the subsequent predication. We can see that they are related *asymmetrically*, since the former serves as a background for the latter, and their order cannot be reversed without changing the meaning of the utterance.

The second criterion relates to the level of language items across which the relations hold. Thus, the same major types of relations exemplified in 1–3 can also hold across propositions, exemplified in examples 4–6, resulting in more complexity within the same types of relations.

Asymmetrical relations across propositions are prototypically signaled by syntactically subordinate constructions, as in (6), where the asymmetric temporal relations between two propositions are manifested explicitly by the subordinating conjunction and different tense agreement in the English translation.

If categorization of discourse relations in terms of complexity is significant apart from purely descriptive considerations, it should prove relevant in areas such as language development, both synchronically (language acquisition) and diachronically (language change). In both areas, there is substantial supporting evidence for the increasing complexity of these relations.

Increasing Complexity in Language Ontogeny, Diachrony, and Typology

The increasing complexity of discourse relations is reflected in the order of development of its markers, both in ontogeny (language acquisition) and in diachrony (e.g., grammaticalization, pidgin and creole studies). Such general trends are reported in the works of Bloom (1973), Bowerman (1973), Bates (1976), Scollon (1976), Givón (1979, 1987, 2009a), Bickerton (1981), Bickerton (1990), Heine and Kuteva (2007), or Sanders et al. (2009), and others.

In stages of child language development, increasing complexity is well documented (Bloom, 1973; Bowerman, 1973; Scollon, 1976; Ochs and Schieffelin, 1979; Givón, 2009b). The focus has typically been on the use of connectives such as conjunctions reflecting age-related command of

² Discourse segments can be characterized as formally distinct units of discourse.

³ Additive relations are also known as iterative (Givón, 2009b).

complex syntax (e.g., Jisa, 1987; McCabe and Peterson, 1991; Berman, 1996; Akinci and Jisa, 2000; Diessel, 2004), where syntactic development progresses from linear juxtaposition. Relations between clauses are first implicit (e.g., Hobbs, 1979; Diessel, 2004) then coordination is acquired, followed by subordination and embedding (Verhoeven et al., 2002). This progression corresponds to additive, symmetrical, and asymmetrical relations, within and across propositions in the Type and Level Approach adopted in this study.

Similarly, in historical change, in both established languages and in creole genesis, the development from additive relations to asymmetric relations across propositions has gained extensive empirical support (e.g., Heine and Kuteva, 2007; Hilpert and Koops, 2009; Koops and Hilpert, 2009). In addition, connective devices initially used at the level of the proposition often grammaticalize into various types of connectors across propositions. For example, the clause-final locative demonstrative *ia* 'here,' originally characterizing a location in a single proposition in Tok Pisin, developed to signal particular relations across propositions – between relative and main clauses (Sankoff and Brown, 1976).⁴

Conventionalization, Convergence, and Reduction

A critical part of language emergence and change is conventionalization. Through conventionalization, a language community converges on a consensus about the relationship between forms and their meanings. In a comparison of two young sign languages, ABSL and ISL, Meir and Sandler (in press) showed that language begins with extensive variation at all levels of structure, before gradually converging on conventionalized forms. They also showed that different parts of the grammar (e.g., phonology and different kinds of morphology) conventionalize at different paces, so that different aspects of linguistic organization should be evaluated in their own right, a direction that we take here.

In established languages, diachronic changes are sometimes driven by a tendency to distinguish grammatical meanings with distinctive forms, to make language more precise and, as a result, more explicit. For example, the Latin subordinating concessive conjunction *quamvis* started out as a clause-internal marker of speech-situation evocation meaning something like 'as you want' and grammaticized to a marker of concessive, adversative relations between propositions. As often happens in grammaticalization, at the beginning of this process, *quamvis* occurred interchangeably with other conjunctions, but gradually became the most frequent marker of the concessive/adversative relation in Latin, interpreted as 'although,' while the usage of other variants for this function decreased drastically. We will report a similar process of convergence and conventionalization in ISL.

⁴Sociolinguistic factors can affect the degree of language complexity in young and emerging languages. This is mostly as a result of external factors, such as the degree of contact with other languages, the size and social structure of the language population, and the number of second language users (Bolender, 2007; McWhorter, 2007; Wray and Grace, 2007; Lupyan and Dale, 2010; Meir et al., 2012; Trudgill, 2012; Meir and Sandler, in press in signed and spoken language research). The comparison of ISL with other languages using the same criteria is a topic for future research.

In the process of conventionalization and specialization, the form of *quamvis* was also phonetically reduced – another measure of language emergence and conventionalization. The principle of economy (reduction of effort in the production of form) constantly interacts with the frequency of a language item, as demonstrated for example by Bybee (2010, 2013) in a series of studies on spoken languages. It is well known that the most frequent expressions tend to be reduced phonetically (see Zipf, 1935). In other words, information that is redundant because it is recoverable and/or predictable, either due to frequency of use or grammatical redundancy, tends to be reduced.

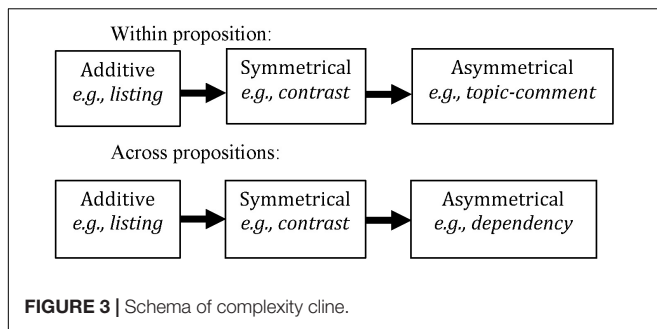
In sum, increase in complexity is accompanied by the development of linguistic signals *specialized* for a particular discourse function in a *systematic* (conventionalized) way, as well as by the *reduction of articulatory effort* (Lehmann, 2008; Bybee, 2010). It should be emphasized that language changes do not happen overnight; old forms do not give way to new without oscillation or variation. Changes occur gradually, as "orderly heterogeneity" is a fundamental characteristic of language (e.g., Weinreich et al., 1968; Labov, 1981). For example, in Labov's (1981) discussion of the *a* split (i.e., tense/lax) in Philadelphia there is a steady movement from 0% tense *a* in the oldest speakers, a slight tendency toward tensing in speakers 40–60 years old, about 30% tensing among speakers in their twenties and thirties, and almost 50% among pre-adolescents and adolescents. Moreover, this growth in the tensing pattern does not occur evenly across the gamut of lexical items, but rather progresses incrementally by particular lexical items. Similar patterns of change have been observed for sign languages (Meir and Sandler, in press).

Research Questions and Hypotheses of the Present Study

The present study elaborates and further supports the Type and Level Approach to understanding linguistic complexity and its emergence. We extend the approach to a language in a different physical modality from that of spoken language – a sign language. As ISL is a young language and some of its earliest users can still be recorded, we can track the development of complexity over time. We take advantage of the unique relation in sign languages between bodily articulations and linguistic forms, to investigate the accumulation of more complexity, more convergence in form, accompanied by reduction in this young language.

We hypothesize that the Type and Level Approach predicts the course of emergence of linguistic complexity in ISL – from more basic additive relations to the more complex asymmetric relations, and from the lower, within-proposition discourse level to a higher, across-proposition level, as schematized in **Figure 3** below.

In addition to increased complexity, we hypothesize that markings will become more systematic (i.e., more frequently used in the expected context), that signers will converge on fewer types of marking for a particular relation (like *quamvis* in Latin described above), and that their form will be reduced as the language matures.



METHODOLOGY

This project follows the approach introduced in Sandler (2012), by analyzing language from the outside in, paying particular attention to the bodily articulators in sign languages and evaluating the linguistic structures that they manifest according to their meaning and discourse relations. Here we present evidence from ISL, a young sign language (about 90 years old) for the gradual emergence of linguistic complexity – from basic to complex language forms – through the gradual increase in systematicity and specificity of bodily signals.

To this end, we investigate the emergence of three relations, varying in terms of relation type (additive, symmetric, and asymmetric) and level (within or across propositions). We compare the frequency and systematicity of each relation produced by fifteen signers of different ages in 2 min each of spontaneous narrative. While spontaneous data are less controlled than elicited data, spontaneous data have the advantage of being natural and more ecologically sound than elicited data in terms of information structuring and the relations among constituents.

We adopt the apparent time hypothesis (Labov, 1963; Sankoff, 2006), which holds that language users do not change their language significantly after young adulthood, so that the language of older generations is a reflection of the state of the language in their youth. A comparison between older and younger signers enables us to make inferences about language emergence from the outset, since the language analyzed in this study is less than 100 years old.

According to the Type and Level Approach, we expect additive constructions to be present in the earliest stages of sign language emergence, as discussed in **Research Questions and Hypotheses of the Present Study**. Therefore, we expect signers of all ages to show similar frequency rates for additive constructions. Conversely, since we expect constructions with more complex relation types, such as symmetric and asymmetric relations, to appear in later stages of language emergence, we predict that younger ISL signers will show more examples of these relations than older ISL signers. Assuming that relations across propositions are more complex than relations within propositions, we expect older signers to convey symmetric and asymmetric relations within propositions more often than across propositions, and younger signers are expected to mark more across proposition relations than older signers.

Participants

Fifteen deaf ISL signers were filmed as part of this study. They represent three age groups, five in each: younger (18–29), middle (30–54), and older (55+ years). Participants ranged in age from 18 to 68 years (mean age: 42 years, 6:9 male: female), and their preferred language was ISL.

The oldest group of signers in our study have varied language backgrounds. Some were born outside of Israel, immigrating to Israel at a young age. Despite this, the first and most preferred language for the older signers in our dataset was ISL. The language of this group, like subsequent age groups, is by no means a home sign system. It qualifies fully as a language, since it has a large, conventionalized vocabulary and linguistic organization; it is the preferred language of its users; and it fulfills all communication and social functions of language throughout the larger community of 10,000 (See Meir and Sandler, 2008 for details). We did not control for heterogeneity (i.e., variation in terms of education and literacy), because it is this sort of variation that characterizes the language of this age group, and which was the model for the younger groups. Younger signers in our dataset are less heterogeneous than older signers, as they were exposed to peers and adult models from a young age, and attended school in deaf education frameworks.

All participants of all three age groups either grew up in deaf families (70% participants in the two younger groups) or have signed from a very young age, and all use sign language as their primary means of communication. This was confirmed by a detailed questionnaire containing information about language use throughout the lifetime of the participants. Consent was obtained from all participants for their involvement in the filming, and signers were compensated for their time. Filming took place at the University of Haifa Sign Language Research Lab.

Task Procedure

Participants were asked to tell a personal life story to a deaf native signer research assistant of middle age. Narratives ranged in length from 3–40 min. We extracted 2 min from the middle of each narrative for this study (30 min of data in total). We did not analyze the beginnings of narratives as our aim was to analyze naturalistic signing, when the signer had become accustomed to the presence of the camera.

Coding Units of Analysis and Discourse Relations

Narratives were divided into intonational phrases based on manual signals. Previous research has demonstrated that manual signals, such as pause, hold and reduplication, correspond to phrase-final lengthening in ISL, and are reliable signals of intonational phrase boundaries, often accompanied by blinks in ISL⁵ (Nespor and Sandler, 1999). The intonational phrase is the main domain of non-manual intonational contours. In other words, facial expression and head movements, corresponding to intonation, systematically align with intonational phrase boundaries (Nespor and Sandler, 1999; Dachkovsky et al., 2013).

⁵Eyeblinks were also found to align with intonational phrase boundaries by Wilbur (1994).

As in spoken languages (e.g., Chafe, 1984; Du Bois, 1985), intonational phrases can be considered as roughly corresponding to thought units. In this study the intonational phrase was the basic unit in the analysis of the distribution of non-manual signals marking discourse relations. All coding was completed using ELAN (Crasborn and Sloetjes, 2008), a video annotation software.

Each discourse relation was identified based on reliable markers in the sign language literature (see **Table 2**). The first relation, additive, is often described in the sign language literature as listing. It is expressed by a movement of the head, often in a thrusting action, along the forward-back axis, with or without movement of the torso. This has been noted in a number of sign languages (Wilbur and Patschke, 1998 in ASL; van der Kooij et al., 2006 in NGT; Sandler, 2012 in ABSL; Puupponen, 2017 in Finnish Sign Language). The second relation, the symmetric relation, is marked by opposite torso or head leans (van der Kooij et al., 2006 in NGT; Man, 2008 in Hong Kong Sign Language; Crasborn and van der Kooij, 2013; Puupponen, 2017 in Finnish Sign Language). A similar contrastive display has been noticed for ISL (Meir and Sandler, 2008).

Finally, a number of articulations of the body have been associated with asymmetric relations. For example, raised brows has been shown to characterize subordinate constructions, such as conditionals and temporal clauses, as well as asymmetric relations within propositions such as between topic and comment (Liddell, 1980; Cecchetto et al., 2006; Dachkovsky and Sandler, 2009; Gökgöz, 2009; Fenlon, 2010). These relations are by definition asymmetrical because one constituent provides background information for the other constituent. Another marker, forward torso lean and/or forward head movement, marks asymmetric relations across the proposition, e.g., in conditional and relative clause constructions (Dachkovsky and Sandler, 2009; Dachkovsky et al., 2013; Dachkovsky, 2017). Forward head movement commonly combines with raised brows to mark dependent constructions (Nespor and Sandler, 1999; Dachkovsky and Sandler, 2009; Dachkovsky et al., 2013).

In total, for the present study we coded 17 articulations. They included movements of the torso (thrust, forward, back, tilt left, tilt right, turn left, and turn right), and head (thrust, forward, back, tilt left, tilt right, turn left, turn right, up, and down), as well as eyebrow raises. We made the distinction between head or torso thrusts or head and torso movements. The two share the same direction, forward and back, while a thrust has a quicker movement.

Data Analysis

From here on, we use the term ‘marking’ to refer to the articulations of the body, which accompany a given construction. We are aware that this term is traditionally restricted to conventionalized grammatical units such as morphemes, but here we follow the sign language literature which has long studied ‘non-manual markers,’ and we use the term in a broader sense, to mean any articulatory action corresponding to a linguistic unit, agnostic with respect to the degree of conventionalization. Another caveat before we proceed: apart from brow raises, our

coding did not include facial expressions⁶, and this awaits future research.

We investigate the frequency in the marking of three different types of relations: additive, symmetric, asymmetric. We follow these relations at two levels of discourse: within and across propositions. Together with discourse relations, we investigate three parameters that accompany the increase of language complexity, as described in **Measuring Relations and Complexity in Discourse**: (1) systematicity, (2) convergence on a particular marker, and (3) articulatory reduction of these markers, which is also an indication of their conventionalization. We measure (1) systematicity in terms of frequency of marking of these relations – we calculate the proportion of intonational phrases marked with each relation during 2 min of narrative and (2) convergence in terms of number of marker variants for the same function. Finally, we measure (3) the reduction of articulatory effort by (a) number of articulators and (b) type of articulators, where a decline in the number of articulators as well as a reduction in size of articulators is an indicator of a reduction in marking. For example, the head is a smaller articulator and its activation results in less muscular activity and displacement in space than the torso. Each specific marking (e.g., torso thrust) was calculated as a fraction of the total marked instances for each relation (e.g., 20% of additive relations marked by a torso thrust). The results of these analyses were further correlated with the complexity of discourse relations in order to see if the changes in the parameters of systematicity, convergence and reduction are indicative of an increase in complexity.

A total of 1473 tokens were analyzed. For every intonational phrase, we counted whether each relation was present, in order to make the tokens binary. For our study, we were interested in whether the frequency of each relation is predicted by the signer’s age. To this end, we included age as an independent and continuous variable (i.e., entering individual ages in years rather than categorical age groups). We conducted multivariate logistic regressions using a program known as Rbrul (Johnson, 2009) for each individual relation. Rbrul quantitatively evaluates the influence of multiple factors on variation (Rand and Sankoff, 1990). We included participant as a random effect in order to account for the effects of individual differences (Baayen, 2008; Jaeger, 2008).

RESULTS





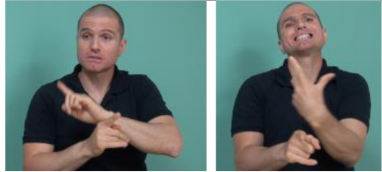
Relations per Intonational Phrase Unit

The total number of relations produced in our dataset was 755. On average, each signer produced 50 relations of varying degrees of complexity during 2 min of spontaneous narrative. We also found that on average signers across age groups produced similar numbers of intonational phrases during 2 min of narrative ($O = 100$, $M = 95$, $Y = 100$).

Table 3 presents the average percentage of all intonational phrases containing each relation, displayed by each age group.

⁶For a study of the emergence of a particular structure in ISL that explicitly incorporated facial expression, see Dachkovsky (2017).

TABLE 2 | Marking for each relation in ISL.

Relation	Gloss of ISL example and translation	Marking in ISL	Example of ISL marking
Additive within proposition (no examples of additive across proposition)	[<u>FATHER</u>] [<u>HAMMER</u>] [SAW] [PAINT] <i>My father was hammering, sawing, and painting.</i>	Head thrusts	
Symmetric within proposition	[<u>TEL-AVIV</u>] [<u>HAIFA</u>] [COMPETE] <i>It was Tel Aviv vs. Haifa.</i>	Contrasting head and/or torso positions	
Symmetric across proposition	[<u>ALSO</u> LANGUAGE SIGN-LANGUAGE] [<u>ALSO</u> ME NO RULE-OUT IMPORTANT ORAL] <i>As well as sign language (being important), I don't rule out the importance of spoken language.</i>	Contrasting head and/or torso positions	
Asymmetric within proposition	[TWO-OF-THEM WIFE <u>FATHER</u>] [TWO-OF-THEM STAY] <i>My father and his wife they both had to stay.</i>	Head forward, brow raise, and retraction	
Asymmetric across proposition	[BECAUSE B-12 <u>NO</u>] [NEED INJECTION] <i>(Doctor said) ...because you're not getting vitamin B12, you need injections.</i>	Head forward, brow raise and retraction	

The underlined words are the signs represented in the images.

Some relations appear in similar distribution across all age groups. For example, additive constructions are present in 0.12 for younger, 0.13 for middle and 0.10 for older signers. Other constructions appear to be more frequent for some age groups than for others. For example, asymmetric relations across

propositions (e.g., clausal dependencies) are more frequent in the narratives of younger signers (0.08) compared to older signers (0.01). We now present our statistical findings for types and levels of relation, leading us to an interpretation of the data relative to our hypotheses.

TABLE 3 | Proportion of relations per intonational phrase.

Construction	Older	Middle	Younger	Total
Additive	0.098	0.13	0.118	0.346
Symmetric within proposition	0.018	0.037	0.042	0.097
Symmetric across propositions	0.03	0.061	0.108	0.199
Symmetry	0.146	0.228	0.268	0.642
Asymmetric within proposition	0.116	0.12	0.074	0.31
Asymmetric across propositions	0.018	0.04	0.076	0.134
Asymmetry	0.134	0.16	0.15	0.444
Total number of intonational phrases	500	475	500	1475

Types of Relations

We investigated age-related differences for additive, symmetric and asymmetric constructions, using statistical methods, to see whether the presence of a particular relation type is predictable by a signer's age. The simple, additive type of construction occurred at the within-proposition level only, and was common across age groups. Multiple regression analysis reveals no significant differences with age for additive relations ($p > 0.05$). However, presence of symmetric constructions (log-odds⁷ -0.029 , $p = 0.015$) and asymmetric constructions (log-odds -0.014 , $p = 0.0233$) was significantly predicted by age. The results show that these relations are used significantly more by younger signers. The results are plotted as percentages in **Figure 4** below.

Levels of Relations

For levels of relations (within or across propositions), we investigated whether there were any age-related differences. Here we found a difference between the two levels. The data show similar numbers of symmetric and asymmetric relations within propositions across all age groups – there were no significant age-related results for symmetric or asymmetric relations within propositions. While no significant age differences were found for symmetric constructions across propositions ($p > 0.05$), we did observe a trend, such that younger signers mark symmetric constructions across propositions more than older signers ($Y = 11\%$, $>M = 6\%$, $>O = 4\%$). We expect this result to be significant with the addition of more tokens.

For asymmetric constructions across propositions (e.g., main/subordinate constructions, i.e., embedding), age was found to be statistically significant. Multiple regression revealed that presence of these relations was predicted by age (log-odds -0.043 , $p < 0.01$), with significantly more relations found in the signing of younger signers.

⁷Log odds measure the strength of the relationship between a factor (in this case, age) and a dependent variable (in this case, presence of a relation type in each intonational phrase). Results with positive log odds, above 0, indicate a positive correlation and results with negative log-odds, below 0, indicate a negative correlation between the variables - the higher the value the stronger the correlation.

Convergence in Marking of Relations

Our results point to increasing convergence; that is, there is less variation in the ways in which relations are marked in the younger signers. The findings for each relation are as follows:

- For the marking of *additive relations*, older signers mark these relations with a number of coexisting variants, with similar frequencies (torso thrust 30%, head and torso thrust 19%, head thrust 25%, etc.) Younger signers employ the same variants as the older signers, but they show clear convergence in the direction of one variant for the marking of additive relations: head thrust 41%.
- For *asymmetric relations within* the proposition, the main marking used by older signers is manual timing, i.e., phrase final lengthening (32%), followed by a number of other frequent variants (brow raise and forward head movement 18%, brow raise alone 15%, and forward head movement alone 12%). Younger signers, however, converge toward two main signals making up the majority of the marking (brow raise and forward head movement 35% or forward head movement alone 38%).
- While 44% of *asymmetric relations across* propositions are marked with a brow raise and forward head movement by older signers, the majority of the marking (56%) is randomly divided among other unrelated signals represented by single occurrences (e.g., head tilt, head turn, or brow lower). Younger signers again show more efficient and systematic use of the body, using only two major variants in equal distribution, one a reduced version of the other (brow raise and forward head movement 34% or forward head movement alone 34%).
- For *symmetric relations within* propositions, older signers use three distinct variants (torso and head tilt 50%, torso tilt 25%, and head tilt 25%) while all of the marking for younger signers is attributed to two variants, i.e., articulation by one articulator or the other (head tilt 53% and torso tilt 47%).
- For *symmetric relations across* propositions, we find no clear reduction in the number of variants used across age groups. That is, we do not see convergence on particular variants. Instead, we see a reduction in the number and type of articulators used (reported in **Changes in Number and Type of Articulators** below).

Changes in Number and Type of Articulators

Further analyses of the data reveals differences in the number and type of articulator activated. We find three main age differences; the marking of relations by younger signers is characterized by: (a) fewer articulators, (b) less use of the torso, and (c) composites are split into individual markers. We present the findings below:

- Fewer articulators – we collapsed our findings to consider the average number of articulators used to mark a discourse relation by signer age group (see **Table 4**). Younger signers in all relation types and levels are more likely to use a single

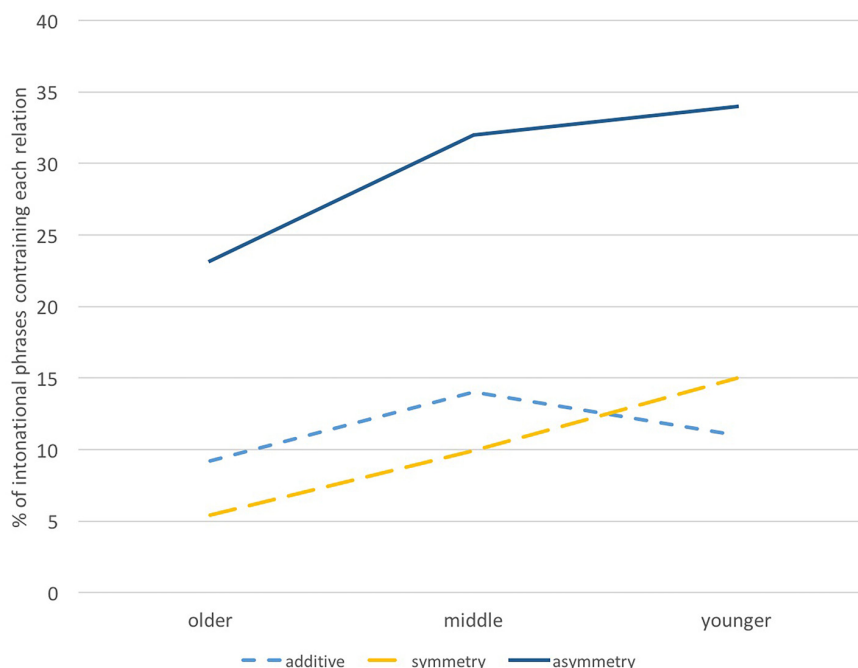


FIGURE 4 | Percentage of intonational phrases containing each relation, displayed by age group.

TABLE 4 | Average percentage of articulators per discourse relation.

Age group	1 Articulator	2 Articulators	3 Articulators	Manual timing
Older	38.6%	45.2%	7.4%	6.6%
Middle	47.2%	45%	5.2%	2.6
Younger	63.8%	33.8%	2.4%	0%

articulator to mark a relation (64%) than older signers (39%). Older signers favor the use of two articulators in most cases.

- (b) Less use of the torso – in additive relations and symmetric relations across propositions we see a reduction in the use of the torso across age groups, with younger signers using the torso less than older signers. Asymmetric relations, within and across propositions, rarely use the movement of the torso and therefore we do not see any change across the age groups for this parameter.
- (c) Composites split into individual markers – in nearly all relations, older signers tend to use a composite of articulators simultaneously for the same relation (e.g., torso and head tilt simultaneously for symmetric within proposition). Younger signers, rather than simply moving toward the use of one dominant marker exclusively, split the composite into individual articulators and use either one of the articulations used in the composite (e.g., torso tilt or head tilt for symmetric within proposition). We discuss the implications of this trend in the **Discussion: Bodily Marking Emerges Gradually**. In **Figure 5** below,

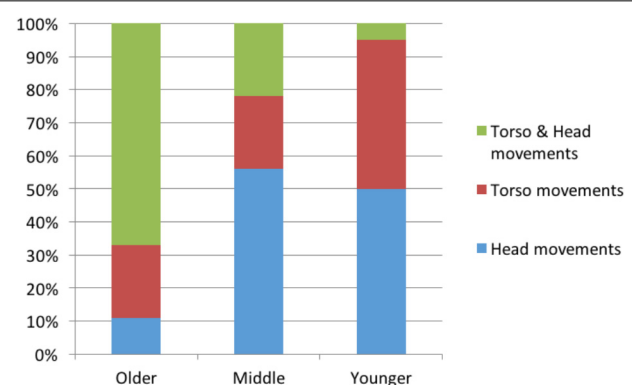
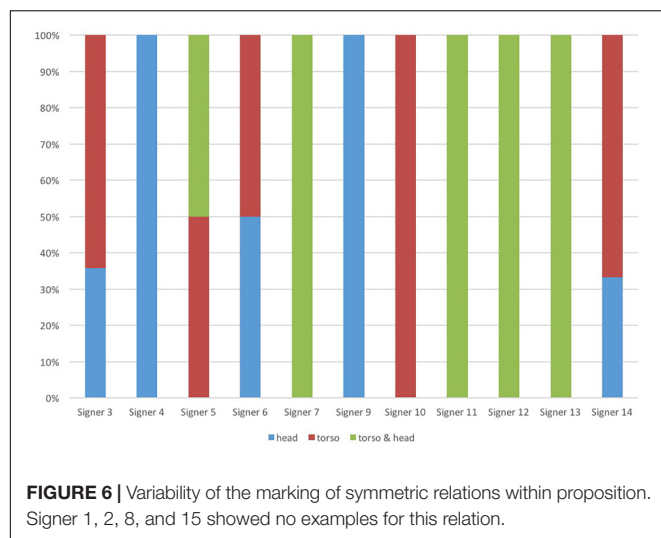


FIGURE 5 | Decrease in multiple marking of symmetric relations within propositions.

we categorize the use of head and torso movements in symmetric relations and demonstrate the reduction in composites and the increase in individual head or torso use (composites: $O = 67\%$, $M = 22\%$, $Y = 5\%$).

While the patterns across age groups are clear, we do find some variation at the level of individual signers, as shown in **Figure 6**. Despite this individual variation, results were significant at the age group level, a finding to which we return below, in **Convergence: Increased Systematicity of Bodily Marking**.



DISCUSSION: BODILY MARKING EMERGES GRADUALLY

Our results suggest that a signer's age – and according to our analysis, the age of the language – is an important predictor of frequency, systematicity, and complexity of relation marking. Frequency of marking varies depending on relation type and level. Our results indicate that younger signers mark more symmetric and asymmetric relations than older signers, despite the fact that all signers produced similar numbers of intonational phrases. There were no significant differences in the frequency of marking for simple additive relations, however. We also found a difference between the frequency of asymmetric relation marking depending on whether they were relations within or across propositions, with younger signers marking more asymmetric relations across propositions than older signers. Systematicity of marking differed across age groups. While older signers used larger articulators (the torso) and often redundantly marked relations with combinations of articulations, younger signers used the torso less, and they used fewer articulators, and often fewer composites, for a given relation. In the next sections, we discuss how our findings may be interpreted in terms of language emergence.

Similarity in Number of Thought Units Across Age Groups

Interestingly, we found no difference in the numbers of intonational phrases produced across age groups. All signers produced roughly 100 intonational phrases during the 2-min narrative. Similar results were found for ABSL (Sandler et al., 2011). The number of signs in an intonational phrase and the internal complexity increased for younger signers, as did the speed of signing, but the number of intonational phrases was constant. In this study, we found that propositions produced by older signers consisted of fewer signs than propositions produced by younger signers. Based on the proposed correspondence between intonational phrases and thought units (Chafe, 1984;

Du Bois, 1985), this finding suggests that humans generally conceive of and express thought units at the same rate, regardless of the internal linguistic complexity of each unit and of their interrelations.

This finding is important in verifying our results as it indicates that older signers are not simply marking fewer relations because they produce fewer or less complex thought units – instead, it strengthens our finding that older signers simply do not have the linguistic means for marking relations. The recruitment of different bodily articulators for different linguistic purposes takes time to emerge in a young sign language.

Simple Adjacency Emerges Before Dependency

We suggest that those relations that are marked statistically more frequently by younger signers are undergoing change; that is, these relations are increasing as the language matures, while those used similarly across age groups are not undergoing change. In terms of language emergence, relations that are no longer undergoing change with age but show a stable use across all age groups (e.g., additive relations, see **Types of Relations**) may reflect earlier constructions that have stabilized in the language development process. Signals which are still undergoing change may reflect constructions that conventionalize later, that take time to emerge in a language. Therefore, since we find no age differences for additive relations, we propose that additive relations emerge before symmetric and asymmetric relations, and that relations within proposition (both symmetric and asymmetric) emerge before relations across propositions (i.e., symmetric and asymmetric across propositions). This supports our hypotheses, as schematized in **Figure 3**. Importantly, this shows that not all relations are marked from the outset of language emergence, and furthermore, that both the type and level of relation are important predictors in the ordering of emergence.

Since the relations that appear in the earliest stages of language emergence – additive type relations and relations within propositions – are less complex, we conclude that simple discourse relations emerge before more complex ones, as hypothesized. This directionality of language development – from additive relations to more dependent relations – has been attested in numerous studies in the framework of grammaticalization, in well-researched European languages, like Latin (Lehmann, 2008), and also in less studied languages, such as Saliba, an Oceanic language (Bril, 2007). Similar findings have been shown in the development from pidgins to creoles (e.g., Bruyn, 1995). However, since all of these studies involve old, well established languages or languages descending from them, conclusions could not be drawn regarding the emergence of a language from scratch. The current study fills this gap.

Convergence: Increased Systematicity of Bodily Marking

In most types of relations that we analyzed, the marking becomes more convergent as the language matures. In other words, signers gradually converge on one or two signals to mark a specific

relation from a larger number of signals. In doing so, the degree of variation decreases over time, with fewer coexisting variants observed in the signing of later generations (i.e., younger signers). In some cases, we see that signers in the older generation fail to explicitly mark the relation at all, only using manual signals to mark the intonational phrase boundary, or mark relations with a number of coexisting variants, with no indication of convergence among older signers toward a particular marker for each relation.

Increase in conventionalization has also been attested in an earlier study on the grammaticalization of relative clause constructions in ISL (Dachkovsky, 2017), as noted above. Similar processes have been attested in child language acquisition in a number of spoken languages (Berman, 1988, 1996; Givón, 2009b; Tomasello and Brooks, 2016). Results of experiments on language acquisition point in the same direction. For example, Hudson Kam and Newport (2009) investigated what learners acquired when their input contained inconsistent grammatical morphemes by manipulating the degree of input inconsistency and the age of the learners (children vs. adult). They demonstrate that only children, not adults, regularize inconsistent input and make their output less variable and more systematic. These results suggest that children may play a unique and important role in creole formation by regularizing grammatical patterns. And indeed, this is the trajectory of changes shown for pidgins and creoles (e.g., Muysken, 2001; McWhorter, 2005; Plag, 2008).

Literature on conventionalization suggests that the effect of regularization through repeated learning and use is amplified more when measured at the level of the whole population, rather than at the level of an individual language user (Smith and Wonnacott, 2010). The study reported here comes to exactly the same conclusions, but the evidence comes from a language in the visual modality. Specifically, we demonstrate that the convergence on specific articulators and increase in systematicity are a cumulative result of comparison across the age groups (see **Figure 5** above), whereas individual signers display considerable variability in their signing, as shown in **Figure 6**.

Reduction of Marking in Terms of Number and Type of Articulators

So far we have demonstrated that as the language matures there is an increase in the frequency and systematicity of marking which directly corresponds to the degree of complexity of these relations. In addition to this, our findings show a gradual reduction in the number and type of articulators, in line with grammaticalization studies in established languages, as outlined below.

Type of Articulators

In addition to the number of articulators, we also find a change in the type of articulator involved in marking. For symmetric relations, older signers typically displace the head and torso together (see **Figure 7**). Younger signers, however, in most cases engage only the head or only the torso in the marking of relations (see **Figure 8**). The former change, from head and torso to only head, shows that the use of a larger, grosser articulator (the torso) is replaced by a smaller, subtler articulator (the head).

A comparable change has been observed in the use of different arm joints – studies on sign language acquisition have found that signers change from first using joints closer to the body (e.g., movement of the shoulder) to using joints further from the body (e.g., movement of the elbow) (Lavoie and Villeneuve, 1999; Takkinen, 2003), causing reduction in the overall size of the sign. This also resembles findings by Mineiro et al. (2017), in which signing size decreased in the early stages of a new sign language. It may be that signers of a new language become gradually more efficient in the use of their articulators and that as a result signing becomes reduced as the signer is able to reduce their production effort. This seems to hold true in studies on young sign languages with later generations of signers using their bodies in a less holistic way compared to earlier generations in the marking of various linguistic functions (Kegl et al., 1999; Aronoff



FIGURE 7 | Marking of a symmetric relation by an older signer with opposite tilts of head and torso together.

head and torso tilt left

head and torso tilt right

EXIST FOOD

NOT-EXIST FOOD

Sometimes there was food and sometimes there wasn't food (The underlined words in the glosses above are the words signed in the corresponding figures.).

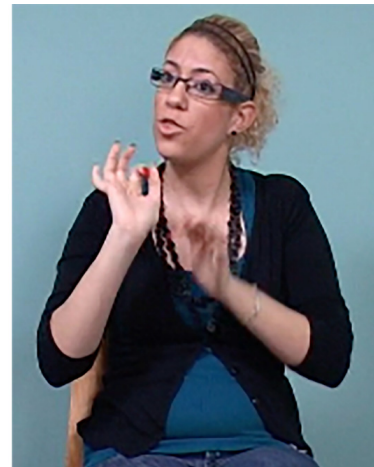
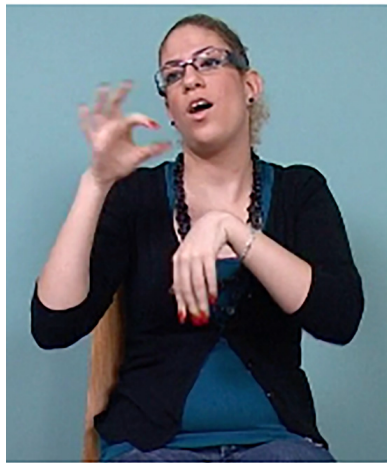


FIGURE 8 | Marking of a symmetric relation by a younger signer with opposite head-only tilts.

head tilt right

head tilt left

BEFORE DOCTOR HE REPLACE

DIFFERENT NEW DOCTOR SHE SIGN GOOD

et al., 2003). Older signers in NSL and signers of ISL (compared to ASL) typically used their whole bodies for representing character viewpoint (i.e., overt constructed action, whole-body classifiers). While reduced effort might account for this replacement of larger articulators (e.g., torso) with smaller articulators (e.g., head), this does not explain why younger signers in ISL continue to use the torso as a relation marker, without moving the head. In the next section, we propose that signers benefit by separating the use of the torso and the head in order for one to be made available to participate in other linguistic functions.

Number of Articulators

Generally, we see a decrease in the number of articulators when we compare across the different age groups. Older signers tend to use multiple articulators – for example, the signer in **Figure 9** moves her head forward, raises her eyebrows and turns her head, compared to younger signers who use a single articulator for the marking of asymmetric relations, such as dependencies (typically, subordination, see **Figure 10**) – for example, with only forward head movement. This change across age groups reflects a diachronic change, such that there is a decrease in the number of articulators used for the marking of these discourse relations as the language matures.

Less Is More: Implications for Compositionality

In later stages of the language, signers use fewer articulators to mark a single linguistic function. What we see here is a reduction in redundancy of articulator marking as the language evolves. During language emergence one might expect that redundancy of feature marking may increase in order to improve comprehensibility (e.g., Pinker and Jackendoff, 2005; Bazzanella,

2011). That said, languages must find a balance between comprehension and economy (Zipf, 1949). ISL at only 90 years of age is considered to be a young sign language⁸, and yet studies of its development have shown that it has changed dramatically relative to other sign languages of a similar age in villages and towns in Israel (Padden et al., 2010a; Meir and Sandler, in press). This has been attributed to the size and heterogeneity of the ISL population, as well as to its use in a range of different domains including education, interpreting programs and the media (Meir et al., 2012). As a result, it is perhaps unsurprising that we see a reduction in redundancy of features in this language, as we find in spoken languages also (e.g., Jaeger, 2010; Bazzanella, 2011). Considering the mixed backgrounds of older deaf signers in our dataset, we might expect to find clearer patterns of reduction in older signers of ABSL, where the sign language community receives relatively little contact from other languages, spoken or signed.

With decreased redundancy and increased systematicity come a number of advantages. In the case of sign languages, by using one fewer articulator in the marking of a specific function, the signer is able to recruit that articulator to mark a different function (Sandler, 2012, 2013 for ABSL). Simultaneous markings of different functions by different articulators do not stand out in our dataset, which we attribute in part to the fact that we restricted our analysis to markers of a subset of discourse relations. For example, markers of information status, indicated by facial expressions (Dachkovsky et al., 2013) and independent actions of the non-dominant hand (Liddell, 2003; Sandler, 2012, under review), were not included in the analysis. We predict that the inclusion of such structures in future research will show

⁸In comparison with British Sign Language, at 300 years old, which is considered to be old among established sign languages (Kyle and Woll, 1985).

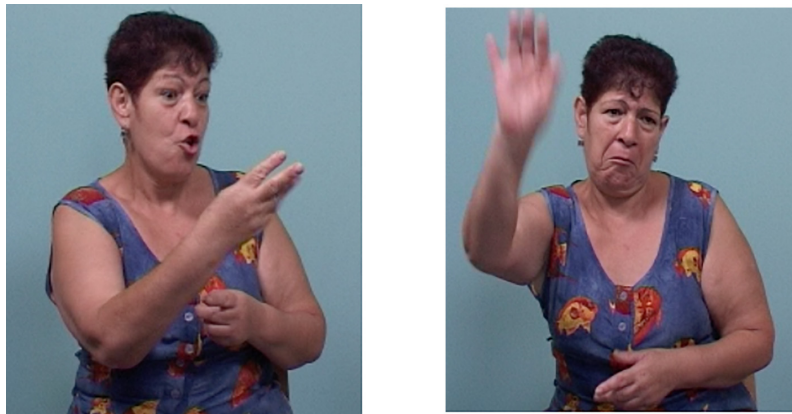


FIGURE 9 | Marking of an asymmetric relation across propositions shown by an older signer – torso shift, forward head movement, head turn, and eyebrow raise accompanies the first part of this construction.

torso and head turn, brow raise, head forward	head tilt left
FATHER SAY YOU TWO <u>THREE</u>	ME ONE <u>GO</u>
<i>My father said that if there had been two or three of us (girls), then I could have gone.</i>	

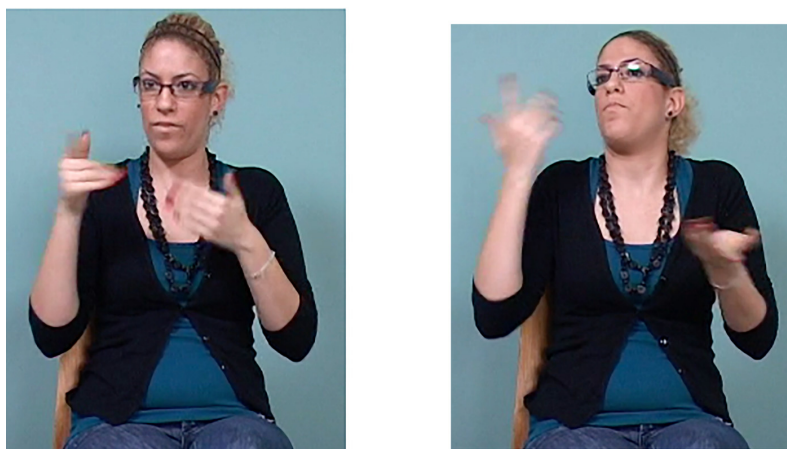


FIGURE 10 | Marking of an asymmetric relation across propositions shown by a younger signer – forward head movement accompanies the first proposition *When my mother needed urgent treatment* and backward head movement accompanies the second proposition, *we did not know what to do*.

head forward	head back
SHE MOTHER EMERGENCY PAST <u>SHE</u>	WE THINK <u>WHAT-TO-DO</u>
<i>When my mother needed urgent treatment we did not know what to do.</i>	

that increased linguistic complexity is reflected by increased simultaneous activation of different articulators for different functions, as demonstrated for ABSL.

However, in the present study, there are some examples of such simultaneous complexity in young signers. For example, one young signer produced an utterance containing two relations (see **Figure 11** above). The utterance can be translated as:

[[[When I was at that school,] [they closed the deaf program]], [and I moved to another school.]]

In this example, the signer marks a symmetric (coordinate) contrastive relation between the two main constituents by tilting his torso to the right for ‘When I was at that school, they closed

the deaf program,’ and to the left for ‘and I moved to another school.’ The first constituent has as its topic ‘that school,’ and topic continuity is marked by keeping the non-dominant hand (‘nd’ – indexing the location of the school) in the signing space throughout.

The information provided in the first constituent is further subdivided into two clauses in an asymmetrical (dependent) relation to one another (‘When I was at that school’ and ‘they closed the deaf program’). Here the signer moves his head forward for the dependent clause, ‘When I was at that school,’ and back for the matrix clause, ‘they closed the deaf program’ – while keeping the body position constant throughout this whole complex first

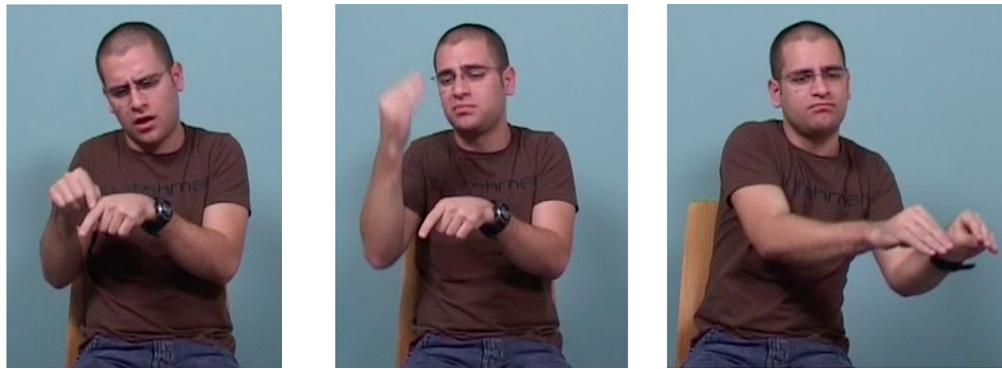


FIGURE 11 | Mapping between simultaneous discourse relations and simultaneous articulations. Opposite torso tilts signal the symmetric contrast between the two major constituents (two different schooling situations); the non-dominant hand (=nd) marks topic continuity; and forward-backward head movement marks the asymmetric relation between dependent and matrix clauses within the first coordinated constituent.

head forward	head back	
torso tilt right		torso tilt left
[[ME GROW-UP SCHOOL THERE] [END SCHOOL THERE CLOSE DEAF INTEGRATE NO-MORE]]		[MOVE-HERE]]
[[When I was at that school,] [they closed the deaf program]], [and I moved to another school.]]		

constituent, and changing it only for the second constituent, ‘*I moved to another school.*’

By separating out articulators for different functions, two discourse relations can be conveyed simultaneously – symmetry by torso tilt and asymmetry by head movement – so that the compositionality of the discourse relations is reflected in the compositionality of bodily articulations.

CONCLUSION

Differences between the frequency of occurrence and the type and consistency of marking of discourse relations by younger and older signers reveal specifically how this young language becomes increasingly complex over time. The most striking finding in this regard is that the asymmetrical relations across propositions – that is, typically, subordination – are significantly more common in the younger than in the older signers. This finding is commensurate with a more limited study on ABSL, where dependent, subordinate structures were found only in younger second generation signers (Sandler et al., 2011).

We also find here that the marking of the discourse relations becomes more systematic over time, in that they become more reliably marked by the same articulators. As the language matures, the signals become more specialized, with fewer articulators dedicated to a particular function, and with finer movements, involving the head more than – and separately from – the torso. Systematic use of the body for linguistic organization mirrors the emergence of linguistic complexity.

The overall picture painted by these results is that of a system that begins with simple relations, unconstrained, redundant form, and high variability. Thus, while the system of older signers clearly has linguistic properties, as we have explained, the aspects

of its organization uncovered here are less systematic. We see gradual change in all of these parameters, resulting in a more conventionalized, systematic, constrained, and compositional sign language.

The previous work on ABSL has suggested that the markers of different discourse functions do not appear all at once. The present study on ISL expands and enriches that proposal by demonstrating that the recruitment of the specific bodily articulators follows a rule-governed functional trajectory – from less complex to more complex discourse functions. Another hypothesis put forward by the earlier study, that the earlier stages of a sign language are characterized by a more holistic use of the body, was also supported in the present study. The current study suggests benefits that the specification of the articulators might contribute in the process of language emergence. Moreover, the findings here shed light on the complex tug of war between conventionalization and regularization on the one hand and variability and diversity on the other (see Meir and Sandler, in press). A lower number of participants in the previous studies might have obscured that issue. In addition, until now, little has been reported at the level of discourse about the path from a system with idiosyncratic characteristics to a more constrained and complex sign language. The work we report here reveals steps along that path.

ETHICS STATEMENT

All participants provided written informed consent to participate in this study. The study was approved by the Ethics Committee/IRB of the University of Haifa (<https://resau.haifa.ac.il/index.php/en/policy-requirement/ethics-committee-institutional-review-board-irb>).

AUTHOR CONTRIBUTIONS

SD is the lead author. RS and SD completed the data coding and analysis. RS contributed toward writing the paper. WS is the principal investigator on the project and contributed toward the writing and interpreting of the results.

FUNDING

This research has received funding from the European Research Council (ERC) under the European Union's Seventh Framework

Programme, grant agreement No. 340140. Principal Investigator, Wendy Sandler.

ACKNOWLEDGMENTS

We wish to thank Shai Davidi for video and other technical assistance, Shiri Barnhart for her administrative help, and Debi Menashe for her ISL expertise. We are grateful to the participants of GestComp 2017 and the 3rd Usage-based Linguistics Conference, Jerusalem for their useful feedback. Special thanks go to all of our deaf participants who were involved in our study.

REFERENCES

- Akinci, M. A., and Jisa, H. (2000). "Development of Turkish clause linkage in the narrative texts of Turkish-French bilingual children in France," in *Studies in Turkish and Turkic Languages: Proceedings of the 9th International Conference on Turkish Linguistics*, (Wiesbaden: Harrassowitz Verlag), 317–324.
- Aronoff, M., Meir, I., Padden, C. A., and Sandler, W. (2008). The roots of linguistic organization in a new language. *Interact. Stud.* 9, 133–153. doi: 10.1075/is.9.1.10aro
- Aronoff, M., Meir, I., Padden, C., and Sandler, W. (2003). "Classifier complexes and morphology in two sign languages," in *Perspectives on Classifiers in Signed Languages*, (Mahwah, NJ: Lawrence Erlbaum Associates), 53–84.
- Baayen, H. R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge, MA: Cambridge University Press. doi: 10.1017/CBO9780511801686
- Baker, C., and Padden, C. A. (1978). *American Sign Language: A Look at Its History, Structure, and Community*. Silver Spring: T.J. Publishers.
- Baker-Shenk, C. (1983). A Micro-Analysis of the Nonmanual Components of Questions in American Sign Language. Silver Spring: T.J. Publishers.
- Barrett, M. (2016). "An introduction to the nature of language and to the themes and issues in the study of language development," in *The Development of Language*, ed. M. Barrett (Hove: Psychology Press), 1–25.
- Bates, E. (1976). *Language and Context: The Acquisition of Pragmatics*. Cambridge, MA: Academic Press.
- Bazzanella, C. (2011). Redundancy, repetition, and intensity in discourse. *Lang. Sci. Discourse Redundancy* 33, 243–254. doi: 10.1016/j.langsci.2010.10.002
- Bergman, B. (1984). "Nonmanual components in signed language: some sentence types in Swedish sign language," in *Recent Research on European Sign Languages*, eds F. Loncke, P. Boyes-Braem, and Y. Lebrun (Lisse: Swets & Zeitlinger), 49–59.
- Berman, R. (1988). On the ability to relate events in narrative. *Discourse Process* 11, 469–497. doi: 10.1080/01638538809544714
- Berman, R. (1996). "Form and function in developing narrative abilities," in *Social Interaction, Social Context and Language*, eds D. I. Slobin, S. M. Ervin-Tripp, J. Gerhardt, A. Kyratzis, and J. Guo (Hillsdale, NJ: Lawrence Erlbaum), 343–367.
- Bickerton, D. (1981). *Roots of Language*. Ann Arbor: Karoma.
- Bickerton, D. (1990). *Language & Species*. Chicago, IL: University of Chicago Press.
- Bloom, L. (1973). *One Word at a Time: The Use of Single Word Utterances before Syntax*. Bloomington: The Hague Mouton Co.
- Bolender, J. (2007). Prehistoric cognition by description: a russellian approach to the upper paleolithic. *Biol. Philos.* 22, 383–399. doi: 10.1007/s10539-006-9058-2
- Bowerman, M. (1973). *Early Syntactic Development: A Cross Linguistic Study with Special Reference to Finnish*. Cambridge, MA: Cambridge University Press.
- Bril, I. (2007). Nexus and juncture types of complex predicates in oceanic languages: functions and semantics. *Lang. Linguist.* 8, 267–310.
- Brown, R. (1973). *A First Language: The Early Stages*. Oxford: Harvard University Press. doi: 10.4159/harvard.9780674732469
- Bruyn, A. (1995). *Grammaticalization in Creoles: The Development of Determiners and Relative Clauses in Sranan [Studies in Language and Language Use 21]*. Ph.D. thesis, Universiteit van Amsterdam, Amsterdam.
- Bybee, J. (ed.) (2010). "Markedness: iconicity, Economy, and Frequency," in *Oxford Handbooks in Linguistic Typology*, (Oxford: Oxford University Press), 132. doi: 10.1093/oxfordhb/9780199281251.013.0008
- Bybee, J. (ed.) (2013). "Usage-Based Theory and Exemplar Representation," in *The Oxford Handbook of Construction Grammar*, (Oxford: Oxford University Press), 49–69.
- Cecchetto, C., Geraci, C., and Zucchi, S. (2006). Strategies of relativization in Italian sign language. *Nat. Lang. Linguist. Theory* 25, 945–975. doi: 10.1007/s11049-006-9001-x
- Chafe, W. (1984). How people use adverbial clauses. *Annu. Meeting Berkeley Linguist. Soc.* 10, 437–449. doi: 10.3765/bls.v10i0.1936
- Coerts, J. A. (1992). *Nonmanual Grammatical Markers. An Analysis of Interrogatives, Negations and Topicalisations in NGT*. Doctoral dissertation, University of Amsterdam, Amsterdam.
- Crasborn, O., and Sloetjes, H. (2008). "Enhanced ELAN functionality for sign language corpora," in *Proceedings of 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora* (ELDA: Paris), 39–43.
- Crasborn, O., and van der Kooij, E. (2013). The Phonology of Focus in sign language of the Netherlands 1 | journal of linguistics | Cambridge core. *J. Linguist.* 49, 515–565. doi: 10.1017/S002226713000054
- Cristofaro, S. (2003). *Subordination*. Oxford: Oxford University Press.
- Cristofaro, S. (2008). "Asymmetric Events, Subordination, and Grammatical Categories," in *Asymmetric Events*, ed. B. Lewandowska-Tomaszczyk (Amsterdam: John Benjamins Publishing Company), 151–172. doi: 10.1075/celcr.11.11cri
- Dachkovsky, S. (2008). "Facial expression an intonation in israeli sign language. The case of neutral and counterfactual Conditionals," in *Signs of the Time: Selected Papers from TISLR 2004*, ed. J. Quer (Hamburg: Signum), 61–82.
- Dachkovsky, S. (2017). *Grammaticalization of Intonation in Israeli Sign Language: From Information Structure to Relative Clause Relations*. Ph.D. dissertation, University of Haifa, Haifa.
- Dachkovsky, S., Healy, C., and Sandler, W. (2013). Visual intonation in two sign languages. *Phonology* 30, 211–252. doi: 10.1017/S0952675713000122
- Dachkovsky, S., and Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Lang. Speech* 52, 287–314. doi: 10.1177/0023830909103175
- Diessel, H. (2004). *The Acquisition of Complex Sentences*. Cambridge, MA: Cambridge University Press. doi: 10.1017/CBO9780511486531
- Dromi, E. (2016). "Early lexical development," in *The Development of Language*, ed. M. Barrett (London: Psychology Press), 99–126.
- Du Bois, J. A. (1985). *Competing Motivations. In Iconicity in Syntax*. Amsterdam: John Benjamins Publishing Company, 343–66. doi: 10.1075/tsl.6.17dub
- Engberg-Pedersen, E. (1990). "Pragmatics of Nonmanual Behaviour in Sign Language," in *SLR '87: Papers from the Fourth International Symposium on Sign Language Research*, (Hamburg: Signum-Press), 121–128.
- Ergin, R. (2017). *Central Taurus Sign Language: A Unique Vantage Point into Language Emergence*. Ph.D. dissertation, Tufts University, Medford, MA.

- Fenlon, J. (2010). *Seeing Sentence Boundaries: The Production and Perception of Visual Markers Signaling Boundaries in Signed Languages*. Ph.D. dissertation, University College London, London.
- Gernsbacher, M. A., and Givón, T. (1995). *Coherence in Spontaneous Text*. Philadelphia, PA: John Benjamins Publishing Company. doi: 10.1075/tsl.31
- Givón, T. (1987). *Beyond Foreground and Background*. In *Coherence and Grounding in Discourse*. Amsterdam: John Benjamins Publishing Company, 175–188. doi: 10.1075/tsl.11.10giv
- Givón, T. (1995). *Functionalism and Grammar*. Philadelphia, PA: John Benjamins Publishing Company. doi: 10.1075/z.74
- Givón, T. (2009a). “Multiple routes to clause union: the diachrony of complex verb phrases,” in *Syntactic Complexity: Diachrony, Acquisition, Neuro-Cognition, Evolution*, eds T. Givón and M. Shibatani (Amsterdam: John Benjamins Publishing Company), 81–118. doi: 10.1075/tsl.85.04mul
- Givón, T. (2009b). *The Genesis of Syntactic Complexity: Diachrony, Ontogeny, Neuro-Cognition, Evolution*. Amsterdam: John Benjamins Publishing. doi: 10.1075/z.146
- Givón, T. (ed.) (1979). “From discourse to syntax: grammar as a processing strategy in discourse and syntax,” in *Syntax and Semantics*, (Michigan: Ann Arbor), 81–112.
- Goddard, C. (2011). *Semantic Analysis: A Practical Introduction*. Oxford: Oxford University Press.
- Gökgöz, K. (2009). *Topics in Turkish Sign Language (Türk Isaret Dili – TİD) Syntax: Verb Movement, Negation, and Clausal Structure*. MSc. Thesis, Boğaziçi University, Istanbul.
- Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help Us Think*. Cambridge, MA: Harvard University Press.
- Graesser, A. C., Millis, K. K., and Zwaan, R. A. (1997). Discourse comprehension. *Annu. Rev. Psychol.* 48, 163–189. doi: 10.1146/annurev.psych.48.1.163
- Heine, B., and Kuteva, T. (2007). *The Genesis of Grammar: A Reconstruction*. Oxford: Oxford University Press.
- Herrmann, A., and Steinbach, M. (eds) (2013). *Nonmanuals in Sign Language*. Vol. 53. *Benjamins Current Topics*. Amsterdam: John Benjamins Publishing Company, doi: 10.1075/bct.53
- Hilpert, M., and Koops, C. (2009). “A quantitative approach to the development of complex predicates: the case of Swedish pseudo-coordination with Sitta ‘sit,’” in *Typological Studies in Language*, Vol. 85, eds T. Givón and M. Shibatani (Amsterdam: John Benjamins Publishing Company), 145–162. doi: 10.1075/tsl.85.06aqu
- Hobbs, J. R. (1979). Coherence and coreference*. *Cogn. Sci.* 3, 67–90. doi: 10.1207/s15516709cog0301.4
- Hudson Kam, C. L., and Newport, E. L. (2009). Getting it right by getting it wrong: when learners change languages. *Cognit. Psychol.* 59, 30–66. doi: 10.1016/j.cogpsych.2009.01.001
- Jaeger, T. F. (2008). Categorical data analysis: away from anovas (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446. doi: 10.1016/j.jml.2007.11.007
- Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognit. Psychol.* 61, 23–62. doi: 10.1016/j.cogpsych.2010.02.002
- Janzen, T., and Shaffer, B. (2002). “Gesture and the Substrate in the Process of ASL Grammaticalization,” in *Modality and Structure in Signed and Spoken Languages*, eds R. P. Meier, R. P. Meier, K. Cormier, and D. Quinto-Pozos (Cambridge: Cambridge University Press).
- Jisa, H. (1987). Sentence connectors in french children’s monologue performance. *J. Pragmat.* 11, 607–621. doi: 10.1016/0378-2166(87)90182-2
- Johnson, D. E. (2009). Getting off the goldvarb standard: introducing Rbrul for mixed-effects variable rule analysis. *Lang. Linguist. Compass* 3, 359–383. doi: 10.1111/j.1749-818X.2008.00108.x
- Kegl, J., Senghas, A., and Coppola, M. (1999). “Creation through contact: sign language emergence and sign language change in Nicaragua,” in *Language Creation and Language Change: Creolization, Diachrony, and Development*, eds M. DeGraff and J. S. Rosenschein (Cambridge, MA: MIT press), 179–237.
- Kendon, A. (1972). “Some relations between body motion and speech,” in *Studies in Diadic Communication*, eds A. W. Siegman and B. Pope (New York, NY: Pergamon), 177–210.
- Kidwell, M. (2013). “Framing, grounding, and coordinating conversational interaction: posture, gaze, facial expression, and movement in space,” in *Body-Language-Communication. An International Handbook on Multimodality in Human Interaction*, ed. C. Müller (Berlin: De Gruyter Mouton), 100–112.
- Kocab, A., Pyers, J., and Senghas, A. (2015). Referential shift in Nicaraguan sign language: a transition from lexical to spatial devices. *Front. Psychol.* 5:1540. doi: 10.3389/fpsyg.2014.01540
- Kocab, A., Senghas, A., and Snedeker, J. (2016). The emergence of temporal language in Nicaraguan sign language. *Cognition* 156, 147–163. doi: 10.1016/j.cognition.2016.08.005
- Koops, C., and Hilpert, M. (2009). “The co-evolution of syntactic and pragmatic complexity: diachronic and cross-linguistic aspects of pseudoclefts,” in *Typological Studies in Language*, Vol. 85, eds T. Givón and M. Shibatani (Amsterdam: John Benjamins Publishing Company), 215–238. doi: 10.1075/tsl.85.09the
- Krifka, M. (2001). For a structured meaning account of questions and answers. *Audiat Vox Sapientia. a Festschrift for Arnim von Stechow*, ed C. Fery & W. Sternefeld, Vol. 52, Berlin: Akademik Verlag, 287–319. doi: 10.1515/9783050080116.287
- Kruijff-Korbayová, I., and Steedman, M. (2003). Discourse and information structure. *J. Logic Lang. Inform.* 12, 249–259. doi: 10.1023/A:1024160025821
- Kyle, J. G., and Woll, B. (1985). *Sign Language: The Study of Deaf People and Their Language*. Cambridge, MA: Cambridge University.
- Labov, W. (1963). The social motivation of a sound change. *Word* 19, 273–309. doi: 10.1080/00437956.1963.11659799
- Labov, W. (1981). Resolving the neogrammarian controversy. *Language* 57, 267–308. doi: 10.2307/413692
- Lackner, A. (2013). *Linguistic Functions of Head and Body Movements in Austrian Sign Language (OGS): A Corpus-Based Analysis*. Ph.D. Thesis, University of Graz, Graz.
- Langacker, R. W. (2009). Constituency, dependency, and conceptual grouping. *Cogn. Linguist.* 8, 1–32. doi: 10.1515/cogl.1997.8.1.1
- Lavoie, C., and Villeneuve, S. (1999). *Acquisition Du Lieu d’articulation En Langue Des Signes Québécoise: Etude de Cas. Variations: Le Langage En Theorie et En Pratique: Le Colloque Des Etudiants et Etudiantes En Sciences Du Langage*. Montreal: University of Quebec at Montreal.
- Lehmann, C. (2008). “Grammaticalization,” in *Theoretical and Empirical Issues in Grammaticalization*, eds E. Seoane and M. J. López-Couso (Amsterdam: John Benjamins Publishing), 207.
- Liddell, S. K. (1978). “An introduction to relative clauses in ASL,” in *Understanding Language through Sign Language Research*, ed. P. Siple (New York, NY: Academic Press), 59–90.
- Liddell, S. K. (1980). *American Sign Language Syntax / Scott K. Liddell. Approaches to Semiotics*, Vol. 52. New York, NY: Mouton.
- Liddell, S. K. (2003). *Grammar, Gesture and Meaning in American Sign Language*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511615054
- Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One* 5:e8559. doi: 10.1371/journal.pone.0008559
- Man, H. C. (2008). *Body Leans in Hong Kong Sign Language: Marking of Contrast*. MA. thesis, The Chinese University of Hong Kong, Hong Kong.
- Mann, W. C., and Thompson, S. A. (1986). *Rhetorical Structure Theory: Description and Construction of Text Structures*. Marina del Rey, CA: Information Sciences Institute.
- Mann, W. C., and Thompson, S. A. (1988). Rhetorical structure theory: toward a functional theory of text organization. *Text* 8, 243–281. doi: 10.1515/text.1.1988.8.3.243
- McCabe, A., and Peterson, C. (1991). *Developing Narrative Structure*. Hillsdale, NJ: Erlbaum.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: University of Chicago Press.
- McWhorter, J. (2005). *Defining Creole*. Oxford: Oxford University Press.
- McWhorter, J. (2007). *Language Interrupted: Signs of Non-Native Acquisition in Standard Language Grammars*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195309805.001.0001
- Meir, I. (1998). Syntactic-semantic interaction in Israeli sign language verbs: the case of backwards verbs. *Sign Lang. Linguist.* 1, 3–33. doi: 10.1075/sll.1.1.03mei

- Meir, I. (2010). "The emergence of argument structure in two new sign languages," in *Syntax, Lexical Semantics and Event Structure*, (Oxford: Oxford University Press), 101–123. doi: 10.1093/acprof:oso/9780199544325.003.0006
- Meir, I., Israel, A., Sandler, W., Padden, C., and Aronoff, M. (2012). The influence of community on language structure: evidence from two young sign languages. *Linguist. Var.* 12, 247–291. doi: 10.1075/lv.12.2.04mei
- Meir, I., and Sandler, W. (2008). *A Language in Space: The Story of Israeli Sign Language*. Didcot: Taylor & Francis.
- Meir, I. and Sandler, W. (in press). "Variation and conventionalization in language emergence: a study of two young sign languages of Israel," in *Linguistic Contact, Continuity and Change in the Genesis of Modern Hebrew*, eds E. Doron, M. Rappaport Hovav, Y. Reshef, and M. Taube (Texas: Benjamins).
- Meir, I., Sandler, W., Ben-Zeev, Y., Lanesman, S., and Sarsur, M. (2016). How does vocabulary evolve in a new language? Insights from research on village sign languages in Israel. *Israel Stud. Lang. Soc.* 8, 11–43.
- Mineiro, A., Carmo, P., Carocha, C., Moita, M., Carvalho, S., Paço, J., et al. (2017). Emerging linguistic features of sao tome and principe sign language. *Sign Lang. Linguist.* 20, 109–128. doi: 10.1075/sll.20.1.04min
- Muysken, P. (2001). The origin of creole languages. *Creol. Contact* 23:157. doi: 10.1075/cil.23.07muy
- Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., and Lee, R. G. (2000). *The Syntax of American Sign Language*. London: The MIT Press.
- Nespor, M., and Sandler, W. (1999). Prosody in Israeli Sign Language. *Lang. Speech* 42, 143–176. doi: 10.1177/00238309990420020201
- Nir, B., and Berman, R. (2010). Complex syntax as a window on contrastive rhetoric. *J. Pragmatics* 42, 744–765. doi: 10.1016/j.pragma.2009.07.006
- Noordman, L., and Vonk, W. (1997). "The different functions of a conjunction in constructing a representation of the discourse," in *Processing Interclausal Relationships in Production and Comprehension of Text*, (Hillsdale, NJ: Lawrence Erlbaum Associates), 75–93.
- Ochs, E., and Schieffelin, B. (1979). *Developmental Pragmatics*. New York, NY: Academic Press.
- Ormel, E., and Crasborn, O. (2012). The prosodic correlates of sentences in signed languages and their relation to language technology: a literature review and suggestions for new types of studies. *Sign Lang. Stud.* 12, 109–145. doi: 10.1353/sls.2011.0019
- Padden, C., Meir, I., Aronoff, M., and Sandler, W. (2010a). "The grammar of space in two new sign languages," in *Sign Languages: A Cambridge Survey*, ed. D. Brentari (New York, NY: Cambridge University Press).
- Padden, C., Meir, I., Sandler, W., and Aronoff, M. (2010b). "Against all expectations: encoding subjects and objects in a new language," in *Hypothesis A/Hypothesis B: Linguistic Explorations in Honor of David M. Perlmutter*, eds D. B. Gerdts, J. C. Moore, and M. Polinsky (Cambridge, MA: MIT press), 383–400.
- Padden, C., Meir, I., Hwang, S. O., Lepic, R., Seegers, R., and Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture* 13, 287–308.
- Pfau, R., and Quer, J. (2010). "Nonmanuals: their prosodic and grammatical roles," in *Sign Languages (Cambridge Language Surveys)*, (Cambridge, MA: Cambridge University Press), 381–402. doi: 10.1017/CBO9780511712203.018
- Pfau, R., Steinbach, M., and Woll, B. (2012). *Sign Language: An International Handbook*. Berlin: Walter de Gruyter. doi: 10.1515/9783110261325
- Pinker, S., and Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition* 95, 201–236. doi: 10.1016/j.cognition.2004.08.004
- Plag, I. (2008). Creoles as interlanguages: syntactic structures. *J. Pidgin Creole Lang.* 23, 307–328. doi: 10.1075/jpcl.23.2.06pla
- Puupponen, A. (2017). The relationship between movements and positions of the head and torso in finish sign language. *Sign Lang. Stud.* 17, 175–214.
- Puupponen, A., Wainio, T., Burger, B., and Janunen, T. (2015). Head movements in finish sign language on the basis of motion capture data. a study of the form and function of nods, nodding, head thrusts, and head pulls. *Sign Lang. Linguist.* 18, 41–89. doi: 10.1075/sll.18.1.02puu
- Rand, D., and Sankoff, D. (1990). *GoldVarb: A Variable Rule Application for the Macintosh*. Montreal: Centre de recherches mathématiques, Université de Montréal.
- Reilly, J. S., McIntire, M., and Bellugi, U. (1990). The acquisition of conditionals in American sign language: grammaticized facial expressions. *Appl. Psycholinguist.* 11, 369–392. doi: 10.1017/S0142716400009632
- Sanders, T. (1997). Semantic and pragmatic sources of coherence: on the categorization of coherence relations in context. *Discourse Process.* 24, 119–147. doi: 10.1080/01638539709545009
- Sanders, T., Spooren, W. P. M., and Noordman, L. (1992). Toward a taxonomy of coherence relations. *Discourse Process.* 15, 1–35. doi: 10.1080/01638539209544800
- Sanders, T., Spooren, W. P. M., and Noordman, L. (1993). Coherence relations in a cognitive theory of discourse representation. *Cogn. Linguist.* 4, 93–133. doi: 10.1515/cogl.1993.4.2.93
- Sanders, T., Spooren, W. P. M., and Noordman, L. G. M. (2009). Toward a taxonomy of coherence relations: discourse processes. *Discourse Process.* 15, 1–35. doi: 10.1080/01638539209544800
- Sandler, W. (2010). Prosody and syntax in sign languages. *Trans. Philol. Soc.* 108, 298–328. doi: 10.1111/j.1467-968X.2010.01242.x
- Sandler, W. (2012). Dedicated gestures in the emergence of sign language. *Gesture* 12, 265–307. doi: 10.1075/gest.12.3.01san
- Sandler, W. (2013). Vive la différence: sign language and spoken language in language evolution. *Lang. Cogn.* 5, 189–203. doi: 10.1515/langcog-2013-0013
- Sandler, W. (2016). *What Comes First in Language Emergence? In Dependencies in Language: On the Casual Ontology of Linguistic Systems*. Berlin: Language Science Press, 67–86.
- Sandler, W. (2018). The body as evidence for the nature of language. *Front. Psychol.* 9:1782. doi: 10.3389/fpsyg.2018.01782
- Sandler, W., Aronoff, M., Padden, C., and Meir, I. (2014). "Language emergence," in *The Cambridge Handbook of Linguistic Anthropology*, eds N. J. Enfield, P. Kockelman, and J. Sidnell (Cambridge, MA: Cambridge University Press), 250–284. doi: 10.1017/CBO9781139342872.012
- Sandler, W., and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge, MA: Cambridge University Press. doi: 10.1017/CBO9781139163910
- Sandler, W., Lillo-Martin, D., Dachkovsky, S., and de Quadros, R. (in press). "Prosody on the hands and face in sign language," in *The Oxford Handbook of Language Prosody*, eds C. Gussenhoven and A. Chen (Oxford: Oxford University Press).
- Sandler, W., Meir, I., Dachkovsky, S., Padden, C., and Aronoff, M. (2011). The emergence of complexity in prosody and syntax. *Lingua. International review of general linguistics. Rev. Int. Linguist. Gen.* 121, 2014–2033. doi: 10.1016/j.lingua.2011.05.007
- Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: systematic structure in a new language. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2661–2665. doi: 10.1073/pnas.0405448102
- Sankoff, G. (2006). "Age: apparent time and real time," in *Elsevier Encyclopedia of Language and Linguistics*, 2 Edn, eds R. E. Asher and J. M. Y. Simpson (Amsterdam: Elsevier Ltd), 412–415.
- Sankoff, G., and Brown, P. (1976). The origins of syntax in discourse: a case study of tok pisin relatives. *Language* 52, 631–666. doi: 10.2307/412723
- Scollon, R. (1976). *Conversations with a One Year Old: A Case Study of the Developmental Foundation of Syntax*. Honolulu: University of Hawaii Press.
- Senghas, A. (1995). *Children's Contribution to the Birth of Nicaraguan Sign Language*. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge.
- Senghas, A., and Coppola, M. (2001). Children creating language: how Nicaraguan sign language acquired a spatial grammar. *Psychol. Sci.* 12, 323–328. doi: 10.1111/1467-9280.00359
- Senghas, A., Coppola, M., Newport, E. L., and Supalla, T. (1997). "Argument structure in Nicaraguan sign language: the emergence of grammatical devices," in *Proceedings of the 21st Annual Boston University Conference on Language Development*, eds E. Hughes, A. Greenhill, and M. Hughes (Boston, MA: Cascadia Press), 550–561.
- Smith, K., and Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition* 116, 444–449. doi: 10.1016/j.cognition.2010.06.004
- Stamp, R., and Sandler, W. (2016). The grammar of the body and the emergence of complexity in sign languages. *Paper presented at Evolang 2016*, New Orleans.

- Sutton-Spence, R., and Woll, B. (1999). *The Linguistics of British Sign Language: An Introduction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139167048
- Takkinen, R. (2003). "Variations of handshape features in the acquisition process," in *Cross-Linguistic Perspectives in Sign Language Research: Selected Papers from TISLR 2000*, eds A. Baker, B. van den Bogaerde, and O. A. Crasborn (Hamburg: Signum), 81–91.
- Tomasello, M., and Brooks, P. (2016). "Early syntactic development: a construction grammar approach," in *The Development of Language*, ed. M. Barrett (London: Psychology Press), 161–191.
- Trudgill, P. (2012). "On the sociolinguistic typology of linguistic complexity loss," in *Language Documentation & Conservation*, Vol. 3, eds F. Seifart, G. Haig, and N. P. Himmelmann (London: Special Publication), 90–95.
- van der Kooij, E., Crasborn, O., and Emmerik, W. (2006). Explaining prosodic body leans in sign language of the Netherlands: pragmatics required. *J. Pragmatics Special Issue* 38, 1598–1614. doi: 10.1016/j.pragma.2005.07.006
- Verhoeven, L., Aparici, M., Cahana-Amitay, D., van Hell, J. G., Kriz, S., and Viguié-Simon, A. (2002). Clause packaging in writing and speech: a cross-linguistic developmental analysis. *Written Lang. Literacy* 5, 135–161. doi: 10.1075/wll.5.2.02ver
- Weinreich, U., Labov, W., and Herzog, M. I. (1968). *Empirical Foundations for a Theory of Language Change*. Texas: University of Texas Press.
- Wilbur, R. (1994). Eyeblinks & ASL phrase structure. *Sign Lang. Stud.* 84, 221–240. doi: 10.1353/sls.1994.0019
- Wilbur, R., and Patschke, C. G. (1998). Body leans and the marking of contrast in American sign language. *J. Pragmatics* 30, 275–303. doi: 10.1016/S0378-2166(98)00003-4
- Wittgenstein, L. (1961). *Tractatus Logico-Philosophicus*. London: Pears & McGuinness.
- Wray, A., and Grace, G. W. (2007). The consequences of talking to strangers: evolutionary corollaries of socio-cultural influences on linguistic form. *Lingua Evol. Lang.* 117, 543–578. doi: 10.1016/j.lingua.2005.05.005
- Zipf, G. (1935). *The Psychobiology of Language: An Introduction to Dynamic Philology*. Cambridge, MA: M.I.T. Press.
- Zipf, G. (1949). Human behavior and the principle of least effort. *J. Clin. Psychol.* 6, 306–573.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Dachkovsky, Stamp and Sandler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Different Approaches to Meaning in Primate Gestural and Vocal Communication

Katja Liebal* and Linda Oña

Comparative Developmental Psychology, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Marianne Gullberg,
Lund University, Sweden

Reviewed by:

Kirsty Emma Graham,
University of York, United Kingdom
Claudio V. Mello,
Oregon Health & Science University,
United States

*Correspondence:

Katja Liebal
katja.liebal@fu-berlin.de

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 12 December 2017

Accepted: 21 March 2018

Published: 05 April 2018

Citation:

Liebal K and Oña L (2018) Different
Approaches to Meaning in Primate
Gestural and Vocal Communication.
Front. Psychol. 9:478.
doi: 10.3389/fpsyg.2018.00478

In searching for the roots of human language, comparative researchers investigate whether precursors to language are already present in our closest relatives, the non-human primates. As the majority of studies into primates' communication use a unimodal approach with focus on one signal type only, researchers investigate very different aspects depending on whether they are interested in vocal, gestural, or facial communication. Here, we focus on two signal types and discuss how meaning is created in the gestural (visual, tactile/auditory) as compared to the vocal modality in non-human primates, to highlight the different research foci across these modalities. First, we briefly describe the defining features of meaning in human language and introduce some debates concerning meaning in non-human communication. Second, with focus on these features, we summarize the current evidence for meaningful communication in gestural as compared to vocal communication and demonstrate that meaning is operationalized very differently by researchers in these two fields. As a result, it is currently not possible to generalize findings across these modalities. Rather than arguing for or against the occurrence of semantic communication in non-human primates, we aim at pointing to gaps of knowledge in studying meaning in our closest relatives, and these gaps might be closed.

Keywords: meaning, vocalization, gesture, modality, referential, intentional, primates, human language

INTRODUCTION

Human language is characterized by a number of 'design features' (Hockett, 1960). *Semanticity*, with signals linked to specific meanings, is closely related to other design features. For example, *arbitrariness* refers to the lack of a natural connection between the signal's signifying form and its signified meaning – the concept to which it refers (de Saussure, 2003/1916). *Duality of patterning* represents the ability to combine a limited set of meaningless components (phonemes) into meaningful structures (morphemes, words) and even longer, more complex sequences (sentences), organized based on specific rules (syntax), while *productivity* refers to the capacity to produce an infinite number of expressions. Since the purpose of such signals is communication, intentional use (*specialization*) is another key feature of language (Hockett, 1960). Although there are various other features characterizing human language, we will focus on this selection, as these features are closely linked to meaning in human language. They are therefore central to our comparative approach to meaning in primate communication.

As comparative researchers assume a gradual evolution of human language, they suggest that precursors to these characterizing features of language are already present in our closest relatives, the non-human primates (hereafter: primates). Consequently, research into their communicative abilities has been shaped by terms and concepts used to characterize human language. It has, however, been questioned whether the use of linguistic terms is justified when describing the communicative abilities of primates (Scott-Phillips, 2008). For example, regarding meaning, Rendall et al. (2009) argue that animal signals merely influence others, but they cannot convey semantic information (but see Scarantino, 2010). While in human language, meaning "...is the central explanatory construct ...[which] arises from the common representational states of speakers and listeners" (Rendall et al., 2009, p. 234), in primates, signalers neither intend to inform others when communicating nor do they consider whether recipients need this information (but see Crockford et al., 2012). Meaningful communication, however, involves "overt intentionality", which requires the sender to "...produce signals with an intention that receivers recognise that the signaller has such intention" (Scott-Phillips, 2016, p. 233). This intentional structure is suggested to be an inherent component of meaningful communication (Grice, 1957), fueling the debate whether such ostensive communication – with the signalers' communicative intention accompanied by their intention to inform – is unique to humans (Scott-Phillips, 2015b) or shared with other primates (Moore, 2016).

Here, we will not argue in favor of or against claims for meaningful communication in other primates. Rather, our aim is to demonstrate that in research on primate communication, meaning is conceptualized very differently depending on the signal type studied. We will contrast research into gestural communication – including visual, but also tactile and auditory gestures – and vocalizations of primates with regard to those features relevant for meaning: the roles of the sender and the recipient for identifying meaning, the relationship between the signal and its reference (*arbitrariness*), intentional use (*specialization*), and the combination of different signals into meaningful sequences (*duality of patterning*, *productivity*). Note that as some of these features are closely related or might partly overlap, it is not always possible to address each of them separately.

Primates communicate via different sensory channels, including olfactory, tactile, visual and auditory signals. However, signals are often not differentiated based on their sensory modality, but are rather categorized based on the different cognitive mechanisms assumed to underlie their usage (Liebal et al., 2013b). Consequently, researchers distinguish gestures, facial expressions, and vocalizations. Here, we focus on gestures and vocalizations only, as there are virtually no studies on meaningful combinations or intentional usage of facial expressions (for exceptions, see Waller et al., 2015; Scheider et al., 2016).

Gestures are movements of the limbs or head (e.g., 'extend arm', 'head shake') or body postures (e.g., 'present'). As they can comprise different sensory modalities, visual gestures (e.g., 'wrist

offer') are differentiated from tactile (e.g., 'slap') and auditory gestures (e.g., 'chest beat'), with the latter producing a sound, which does not engage the vocal folds (Call and Tomasello, 2007). Vocalizations (or calls), however, emerge from the larynx, unlike sounds (e.g., 'raspberries', 'whistles'), which do not engage the focal folds.

RELATIONSHIP(S) BETWEEN SIGNALS AND THEIR REFERENTS (ARBITRARINESS)

Whether signals have meaning(s) is closely linked to whether they refer to specific referents. In human communication, the exact nature of the relationship(s) between a signal and its referent may vary, as reference is differently conceptualized across disciplines and modalities (Leavens et al., 2005). While linguists use "reference" synonymously with symbolic reference to indicate that in spoken language, "a word stands for something", developmental psychologists also consider non-verbal means of communication in the form of referential gestures, such as pointing gestures of pre-linguistic children (Iverson and Goldin-Meadow, 2005), which can be used to refer to different referents. Furthermore, while linguists highlight the arbitrary relationship between a word and its referent, developmental psychologists suggest that for pointing gestures, the triadic relationship between signaler, recipient, and the external entity is not arbitrary, as "...a point's specific meaning is determined in large part by the spatial locations of the pointer, the thing indicated, and the communicative partner" (Leavens et al., 2005, p. 185). Together this shows that in human communication, reference is treated differently in spoken language compared to visual non-verbal communication.

Likewise, for primates, comparative researchers operationalize reference differently depending on signal type. Vocal researchers focus on context-specific vocalizations to find "...the animal equivalent to referential words in human language" (Liebal et al., 2013b, p. 399), in the form of functionally referential vocalizations. They are produced in response to a specific stimulus (the referent, e.g., a predator), with receivers showing a specific response to these calls, even in the absence of the eliciting stimulus, indicating that this response itself is stimulus-independent (Macedonia and Evans, 1993; Evans, 1997). As it is unclear whether primates' calls refer to a specific referent, for example, a predator ("leopard"), or are requesting a specific action in response to this referent ("go up tree"), the term "functional" is used for primates' referential vocalizations to distinguish them from human referential communication. The ground-breaking finding that vervet monkeys use distinct predator-specific alarm calls in encounters with their main predators (eagles, leopards, pythons) sparked great interest in such functionally referential vocalizations, as playback experiments confirmed that the monkeys showed predator-specific responses upon hearing the corresponding alarm call (Seyfarth et al., 1980). Although claims suggesting the "word-like" nature of these alarm calls

have not been confirmed, many following studies found evidence for functionally referential vocalizations in many primate species. They vary in their degree of specificity as they may refer to specific (e.g., leopard versus eagle) or types of predators (e.g., aerial or terrestrial) (Schel et al., 2009), and have been found in different contexts, such as predation, feeding, or social behavior (Di Bitetti, 2003; Slocombe et al., 2010).

In the gestural modality, visual signals in the form of pointing gestures have received considerable attention regarding their referential function. In humans, pointing gestures emerge early in ontogeny (Liszkowski et al., 2004), and are used to refer to different external entities, such as objects, persons, or events. Thus, pointing gestures have no one-to-one referential meaning; instead, the meaning of a pointing gesture depends on its context of use and the common ground shared by the gesturer and the recipient (Liebal et al., 2013a). In primates, the use of pointing gestures has mostly been studied in interactions with humans (Call and Tomasello, 1994; Leavens et al., 1996; but see Vea and Sabater-Pi, 1998; Hobaiter et al., 2014), within which they use these gestures to request food rewards or objects they cannot obtain otherwise (Bullinger et al., 2011). Like in humans, the meaning of primates' points depends on the context and the common ground primates share with the human experimenter (Bohn et al., 2016). Iconic gestures represent another type of referential gestures, which depict specific objects or actions, resulting in a non-arbitrary relationship between the gesture and the referent. Although concepts of iconicity differ across studies (Perlman et al., 2014), there is some evidence that primates use iconic gestures, mostly to request specific actions (Tanner and Byrne, 1996; Douglas and Moscovice, 2015).

Together this shows that the nature of relationships between a signal and its referent(s) varies across modalities: while some vocalizations are functionally referential signals that refer to specific referents, the relationship between a gesture and its referent(s) varies across gesture types. Pointing gestures can be used to refer to different entities, while iconic gestures depict specific actions.

INTENDED VERSUS EXTRACTED MEANING: THE ROLES OF SIGNALERS AND RECIPIENTS (*SEMANTICITY AND SPECIALIZATION*)

Inspired by ethology, some scholars suggest differentiating between the “messages” of the signaler and the “meaning” extracted by the receiver (Smith, 1965; Font and Carazo, 2010). Meaning is thus conceptualized very differently depending on whether the focus is on the signaler's or recipient's behavior. Vocal studies traditionally focus on the recipient. By using playback studies, researchers investigate recipients' responses toward specific vocalizations to extract their meaning, while they consider contextual information or the signaler's

behavior to a much lesser extent than gesture studies. In the gestural domain, it is not possible to use similar playback experiments to elicit responses to specific gestures at least in interactions between conspecifics. Therefore, unlike in vocal communication, gesture researchers focus on the signaler and investigate whether they produce their gestures intentionally. The term “intentional” is applied in a sense that an individual communicates in a purposeful, goal-directed way, by means of voluntarily controlled actions, while this does not necessarily imply that the recipient understands a signaler's gesture as an intended act of communication. It is also debated whether apes who gesture intentionally could additionally be said to act with communicative intentions (Scott-Phillips, 2015a; Moore, 2016; Townsend et al., 2017). Furthermore, unlike in vocal studies, gesture research largely ignores context-specific signals, as flexible usage is an important criterion to identify intentional communication. Therefore, researchers focus on those gestures used across different contexts and argue that the meaning of a gesture might differ depending on the context in which it is used (Call and Tomasello, 2007). However, although contextual information contributes to identifying a gesture's meaning, “context” should not be used as a substitute for “meaning”. More recently, gesture researchers have started to also consider recipients' responses to identify the signalers' intended meaning when performing a gesture (more in section “New Developments and the Way Forward in Studying Meaning in Primate Communication”), which is more in keeping with vocal research. Importantly, note that “intentional gesture production” has to be distinguished from the “signaler's intended meaning”: A message is only taken to have an intended meaning (as distinct from an intended effect) if it was produced not only intentionally, but with communicative intent – that is, if it was produced both intentionally and ostensively (Scott-Phillips, 2015a; Moore, 2016).

Unlike in the gestural modality, intentionality in vocal production has received little attention. Vocalizations have been suggested to be involuntary expressions of emotional states (Tomasello, 2008), supported by neurobiological studies indicating that vocal production is largely mediated by several motor nuclei in the pons and the reticular formation in the medulla, with no direct connections to cortical motor areas (Jürgens, 2002). This traditional notion, however, is increasingly being challenged, as it has been shown that several cortical areas (e.g., anterior cingulate gyrus and ventrolateral prefrontal cortex) are involved in the production of volitional calls (Gavrilov et al., 2017). Furthermore, as vocal researchers have started to consider the signaler's behavior, they found that chimpanzees' alarm calls are most likely intentionally produced signals (Schel et al., 2013). Chimpanzees even seem to consider conspecifics' knowledge states, as they only vocalize when unknowledgeable individuals are close to a hidden predator (Crockford et al., 2012).

Thus, to determine a signal's meaning, gesture researchers usually focus on the signaler's behavior, while vocal researchers consider the recipient's reactions. However, research on both

modalities is increasingly investigating both signalers’ and recipients’ behaviors to extract the meaning of vocal and gestural signals.

CREATION OF NEW MEANINGS
(DUALITY OF PATTERNING,
PRODUCTIVITY, SYNTAX)

Duality of patterning and productivity are two design features of human language, which both relate to creating new meaningful utterances from an existing, potentially limited repertoire. Comparative researchers are therefore interested in whether primates also combine their signals into meaningful sequences. They investigate whether combinations of several signals are used for different functions than the components they consist of, or alternatively, whether the meaning of one of the components is modified by the other component. Combining several signals is closely linked to the question of whether a specific order is crucial for the creation of new meaning and thus whether such combinations are based on specific syntactical rules.

Zuberbühler (2002) demonstrated that in some situations, Campbell’s monkeys combine their alarm calls with a preceding boom-call, which modifies the meaning of the following alarm call. Thus, while the functionally referential alarm call is uttered in encounters with their predators, they use this specific call combination in less dangerous situations, such as falling trees. Proceeding from this finding, later studies concluded that “...the Campbell’s monkey call system may be the most complex example of ‘proto-syntax’ in animal communication known to date” (Ouattara et al., 2009). A different system was found in Putty-nosed monkeys that use two alarm calls, which are not predator-specific. Interestingly, the reference to specific predators is achieved by producing sequences of calls, as hack-sequences are more likely to be used in response to eagles, while pyow-sequences occur in response to leopards (Arnold and Zuberbühler, 2006). Combinations of the two call types, however, are used to initiate group travels, indicating that by combining these different vocalizations, new meaning is created.

In the gestural domain, sequences are defined as multiple gestures produced one after the other by one individual, toward the same recipient and the same goal, with sequences varying in the number of gestures combined (Liebal et al., 2004). Although findings across species and studies differ, common conclusion are that gestures are not combined in ways to create new meanings and that gesture combinations are not governed by specific rules (e.g., Genty and Byrne, 2010; Roberts et al., 2013; Hobaiter et al., 2014).

This suggests that gesture combinations are not based on combinatorial rules and are not used for different functions than their single components like it has been shown for vocalizations. However, the finding that primates are able to combine vocalizations into more complex sequences with specific meanings is also debated, as “...there is no evidence of the compositional essential to language—having a few sequences

TABLE 1 | Different approaches to studying meaning in primates’ gestural and vocal communication.

	Gestures	Vocalizations
Who is studied?	Signaler (and recipient)	Recipient
Is intentionality considered?	Yes (production)	No
How is meaning studied?	Intended meaning	Extracted meaning
Which signals are studied?	Context-unspecific	Context-specific
What is the relationship between signal and referent?	No one-to-one referential meaning between pointing gestures and their varying referents	One-to-one referential meaning between functionally referential signals and their referents
Which species are studied?	Apes	Monkeys
Which methods are used?	Observations	Playback experiments
Where are studies conducted?	Captivity	Natural habitats

with a well-defined meaning does not qualify as syntax” (Arbib et al., 2008).

NEW DEVELOPMENTS AND THE WAY
FORWARD IN STUDYING MEANING IN
PRIMATE COMMUNICATION

We have shown that meaning in primates is conceptualized and studied very differently in the gestural as compared to the vocal modality (Table 1). Rather than representing fundamental differences across modalities, this may reflect different research traditions and historical limitations in methodological approaches. While gesture research focuses on signalers and whether they communicate intentionally, vocal researchers study the recipients’ responses to identify the meaning they extract from a call. Gesture researchers highlight the importance of the context in which an interaction takes place, as it contributes to a gesture’s meaning. They focus on flexible gesture usage as an important characteristic of intentional communication, and are less interested in context-specific gestures. Vocal researchers, however, traditionally focus on context-specific, functionally referential vocalizations. Slocombe et al. (2011) further demonstrated that gestures are usually studied in great apes, in captive settings, by using observational methods, while most research on vocalizations is conducted with monkey species, in their natural habitats, by using experimental methods. These fundamental differences in how meaning is studied across modalities hinder comparisons across signal types and make it difficult to conclude whether there is evidence for meaningful communication in primates. Furthermore, it seems that researchers are often not aware that they use the term meaning very differently, which in turn does not support a fruitful discourse about how comparative approaches contribute to our understanding of language evolution (Bar-On and Moore, 2017).

However, in both vocal and gestural research, traditional approaches have been questioned and new approaches for studying meaning have been suggested. For example, in the vocal modality, the concept of functionally referential vocalizations has been recently criticized for a number of

different reasons (see Wheeler and Fischer, 2012; Fischer and Price, 2016). First, because of the strong focus on context-specific vocalizations, the prevalence and significance of functionally referential vocalizations might have been overestimated as compared to other, less context-specific vocalizations. Second, it is often assumed that these vocalizations might require more sophisticated cognitive skills than other vocalizations or other signal types, since the differentiated responses of receivers of such calls "...have been widely interpreted as evidence that signals elicit mental representations in receivers based on the information extracted from the signal" (Wheeler and Fischer, 2012, p. 199). However, such specific responses may be explained by lower-level mechanisms such as classical conditioning, "...without drawing on the concept of information, the meaning of calls, or mental representations of a signal's purported referent in listeners" (Wheeler and Fischer, 2012, p. 199). Because of this, the relationship between a vocal signal and its referent might not be as arbitrary as previously suggested. Wheeler and Fischer (2012) therefore suggested abandoning the concept of functionally referential vocalizations. Rather, meaning in primate vocal communication should be studied in the framework of pragmatics to investigate how primates use contextual information – in addition to the information provided by the signal itself (Wheeler and Fischer, 2012).

In the gestural domain, we can observe the opposite trend. While gesture researchers have previously proposed that gestures do not have inherent meaning, but have rather highlighted the importance of the context for defining the meaning of a gesture (Call and Tomasello, 2007), recent studies emphasize that gestures indeed have specific meaning (Cartmill and Byrne, 2010; Hobaiter and Byrne, 2014; Graham et al., 2018). This new approach focuses on both the signaler and the recipient by investigating if the signaler's intended meaning when using a gesture matches with a particular outcome (Byrne et al., 2017). If the recipient's response satisfied the signaler – evident in the signaler stopping the production of a certain gesture – this is referred to as the "apparently satisfactory outcome" of this specific gesture. In other words, the matching of the intended and extracted meaning is used as an approximation of the gesture's meaning. Hobaiter and Byrne (2014) found that wild chimpanzees use at least some gestures with tight meaning, in a sense that the same outcome was observed in more than 70% of their use, while other gestures have loose meaning, as they elicited the same outcome in only 50–70%. Note, however, that chimpanzees still used the majority of their gestures for multiple outcomes (Hobaiter and Byrne, 2014), as found by other studies, which focused on the flexibility of gesture use, and which therefore concluded that chimpanzee gestures have no inherent meaning, as the meaning is defined by the context in which they are used (Tomasello et al., 1994). This shows that conclusions drawn from such studies depend at least partly on which findings are emphasized: while some authors focus on those gestures flexibly used across different contexts and conclude that gestures have no meaning, others focus on context-specific gestures, used for one or few outcomes, and consequently emphasize their specific meaning(s). Future research should

bring together these two perspectives and study context-specific and unspecific gestures in concert, as gesture types may vary in their degree of specificity, as found for vocalizations. For example, chimpanzees' visual gestures are more likely to occur in a specific context (e.g., sexual behavior, requesting food) and thus represent "intention movements", which are abbreviations of full-fledged behaviors used for a specific purpose (Tomasello et al., 1989), while tactile and auditory gestures are often produced across different contexts to trigger others' actions (Liebal and Call, 2012).

Related to this, it is important to discuss how specific meaning has to be, particularly if we aim at comparing meaning across modalities (Scarantino, 2013). Thus, we have shown that pointing gestures have no one-to-one referential meaning, as pointing can be used to refer to different entities, while functionally referential vocalizations often refer to specific referents. Furthermore, researchers differentiate between tight and loose meanings of chimpanzee gestures, and even gestures with tight meanings may result in multiple outcomes. This highlights the lack of definitions applicable across modalities as well as a lack of a measure based on which the specificity of meaning of a signal can be judged.

Sievers and Gruber (2016) therefore suggest using a "pragmatic notion of reference" that focuses on the use of a signal to refer to something in a specific situation – rather than expecting that signals have referential meaning in themselves. They further highlight that in human language, reference is an action of the signaler and claim that "...any definition describing reference in non-human animals must also focus on the producer" (Sievers and Gruber, 2016, p. 759). In other words, a signal only has referential meaning if the signaler intends to refer to a specific referent. This has important implications, as functionally referential vocalizations have been almost exclusively studied with focus on the recipient. Therefore, to be able to conclude that vocalizations are indeed meaningful, we would have to additionally demonstrate that they are intentionally produced.

Finally, we have argued that unlike in the vocal modality, there is currently no evidence for meaningful combinations in gesture sequences. This may be partly explained by the fact that there is only little research investigating if single gestures are meaningful units. As a result, we are currently lacking sufficient datasets to determine whether a gesture's meaning changes when it is part of a sequence compared to when it is used in isolation. Furthermore, we want to highlight the importance of multimodal approaches (Slocumbe et al., 2010), as it is currently unclear whether combinations consisting of different signal types, such as gesture plus facial expression or gesture plus vocalization result in the creation of new meaning or the modification of an existing one (Hobaiter et al., 2017; Wilke et al., 2017). To study meaningful communication in primates in more comprehensive ways, the essential first step is to combine research efforts across modalities, based on shared definitions which are applicable across signal types and to use a multi-perspective approach, which considers the behavior of both signalers and recipients, in addition to the context.

AUTHOR CONTRIBUTIONS

KL wrote the first draft of this manuscript. Both authors then finished the manuscript together and circulated it several times until the current version was finalized.

FUNDING

This paper was supported by the ERC project “The Grammar of the Body: Revealing the Foundations of Compositionality

REFERENCES

- Arbib, M. A., Liebal, K., and Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Curr. Anthropol.* 49, 1053–1076. doi: 10.1086/593015
- Arnold, K., and Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature* 441:303. doi: 10.1038/441303a
- Bar-On, D., and Moore, R. (2017). “Pragmatic interpretation and signaler-receiver asymmetries in animal communication,” in *The Routledge Handbook of Philosophy of Animal Minds*, eds K. Andrews and J. Beck (London: Routledge), 291–300.
- Bohn, M., Call, J., and Tomasello, M. (2016). The role of past interactions in great apes’ communication about absent entities. *J. Comp. Psychol.* 130, 351–357. doi: 10.1037/com0000042
- Bullinger, A. F., Zimmermann, F., Kaminski, J., and Tomasello, M. (2011). Differential social motives in the gestural communication of chimpanzees and human children. *Dev. Sci.* 14, 58–68. doi: 10.1111/j.1467-7687.2010.00952.x
- Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., and Tanner, J. (2017). Great ape gestures: intentional communication with a rich set of innate signals. *Anim. Cogn.* 20, 755–769. doi: 10.1007/s10071-017-1096-4
- Call, J., and Tomasello, M. (eds) (2007). *The Gestural Communication of Apes and Monkeys*. New York: Lawrence Erlbaum Associates.
- Call, J., and Tomasello, M. (1994). Production and comprehension of referential pointing by orangutans. *J. Comp. Psychol.* 108, 307–317. doi: 10.1037/0735-7036.108.4.307
- Cartmill, E. A., and Byrne, R. W. (2010). Semantics of primate gestures: intentional meanings of orangutan gestures. *Anim. Cogn.* 13, 793–804. doi: 10.1007/s10071-010-0328-7
- Crockford, C., Wittig, R. M., Mundry, R., and Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Curr. Biol.* 22, 142–146. doi: 10.1016/j.cub.2011.11.053
- de Saussure, F. (2003/1916). *Cours de Linguistique Générale*. Paris: Payot & Rivages.
- Di Bitetti, M. S. (2003). Food-associated calls of tufted capuchin monkeys (*Cebus apella nigritus*) are functionally referential signals. *Behaviour* 140, 565–592. doi: 10.1163/156853903322149441
- Douglas, P. H., and Moscovice, L. R. (2015). Pointing and pantomime in wild apes? Female bonobos use referential and iconic gestures to request genito-genital rubbing. *Sci. Rep.* 5:13999. doi: 10.1038/srep13999
- Evans, C. S. (1997). “Referential signals,” in *Perspectives in Ethology*, eds D. H. Owings, M. D. Beecher, and N. S. Thompson (London: Plenum Press), 99–143.
- Fischer, J., and Price, T. (2016). Meaning, intention, and inference in primate vocal communication. *Neurosci. Biobehav. Rev.* 82, 22–31. doi: 10.1016/j.neubiorev.2016.10.014
- Font, E., and Carazo, P. (2010). Animals in translation: why there is meaning (but probably no message) in animal communication. *Anim. Behav.* 80, e1–e6. doi: 10.1016/j.anbehav.2010.05.015
- Gavrilov, N., Hage, S. R., and Nieder, A. (2017). Functional specialization of the primate frontal lobe during cognitive control of vocalizations. *Cell Rep.* 21, 2393–2406. doi: 10.1016/j.celrep.2017.10.107
- Genty, E., and Byrne, R. W. (2010). Why do gorillas make sequences of gestures? *Anim. Cogn.* 13, 287–301. doi: 10.1007/s10071-009-0266-4
- Graham, K. E., Hobaiter, C., Ounsley, J., Furuichi, T., and Byrne, R. W. (2018). Bonobo and chimpanzee gestures overlap extensively in meaning. *PLoS Biol.* 16:e2004825. doi: 10.1371/journal.pbio.2004825
- Grice, H. P. (1957). Meaning. *Philos. Rev.* 64, 377–388. doi: 10.2307/2182440
- Hobaiter, C., and Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Curr. Biol.* 24, 1596–1600. doi: 10.1016/j.cub.2014.05.066
- Hobaiter, C., Byrne, R. W., and Zuberbühler, K. (2017). Wild chimpanzees’ use of single and combined vocal and gestural signals. *Behav. Ecol. Sociobiol.* 71:96. doi: 10.1007/s00265-017-2325-1
- Hobaiter, C., Leavens, D. A., and Byrne, R. W. (2014). Deictic gesturing in wild chimpanzees (*Pan troglodytes*)? Some possible cases. *J. Comp. Psychol.* 128, 82–87. doi: 10.1037/a0033757
- Hockett, C. F. (1960). The origin of speech. *Sci. Am.* 203, 89–96. doi: 10.1038/scientificamerican0960-88
- Iverson, J. M., and Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychol. Sci.* 16, 367–371. doi: 10.1111/j.0956-7976.2005.01542.x
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* 26, 235–258. doi: 10.1016/S0149-7634(01)00068-9
- Leavens, D. A., Hopkins, W. D., and Bard, K. A. (1996). Indexical and referential pointing in chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* 110, 346–353. doi: 10.1037/0735-7036.110.4.346
- Leavens, D. A., Russell, J. L., and Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (*Pan troglodytes*). *Child Dev.* 76, 291–306. doi: 10.1111/j.1467-8624.2005.00845.x
- Liebal, K., and Call, J. (2012). The origins of non-human primates’ manual gestures. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 118–128. doi: 10.1098/rstb.2011.0044
- Liebal, K., Call, J., and Tomasello, M. (2004). Use of gesture sequences in chimpanzees (*Pan troglodytes*). *Am. J. Primatol.* 64, 377–396. doi: 10.1002/ajp.20087
- Liebal, K., Carpenter, M., and Tomasello, M. (2013a). Young children’s understanding of cultural common ground. *Br. J. Dev. Psychol.* 31, 88–96. doi: 10.1111/j.2044-835X.2012.02080.x
- Liebal, K., Waller, B. M., Burrows, A. M., and Slocumbe, K. E. (2013b). *Primate Communication: A Multimodal Approach*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139018111
- Liszkowski, U., Carpenter, M., Henning, A., Striano, T., and Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Dev. Sci.* 7, 297–307. doi: 10.1111/j.1467-7687.2004.00349.x
- Macedonia, J. M., and Evans, C. S. (1993). Essay on contemporary issues in ethology: variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology* 93, 177–197. doi: 10.1111/j.1439-0310.1993.tb00988.x
- Moore, R. (2016). Meaning and ostension in great ape gestural communication. *Anim. Cogn.* 19, 223–231. doi: 10.1007/s10071-015-0905-x
- Ouattara, K., Zuberbühler, K., N’goran, E. K., Gombert, J.-E., and Lemasson, A. (2009). The alarm call system of female Campbell’s monkeys. *Anim. Behav.* 78, 35–44. doi: 10.1016/j.anbehav.2009.03.014
- Perlman, M., Clark, N., and Tanner, J. E. (2014). “Iconicity and ape gesture,” in *Proceedings of the 10th International Conference: The Evolution of Language (EVOLANG 10)*, eds E. Cartmill, S. Roberts, H. Lyn and H. Cornish (Singapore: World Scientific), 228–235.
- Rendall, D., Owren, M. J., and Ryan, M. J. (2009). What do animal signals mean? *Anim. Behav.* 78, 233–240. doi: 10.1016/j.anbehav.2009.06.007

- Roberts, A. I., Vick, S.-J., and Buchanan-Smith, H. M. (2013). Communicative intentions in wild chimpanzees: persistence and elaboration in gestural signalling. *Anim. Cogn.* 16, 187–196. doi: 10.1007/s10071-012-0563-1
- Scarantino, A. (2010). Animal communication between information and influence. *Anim. Behav.* 79, e1–e5. doi: 10.1016/j.anbehav.2010.03.005
- Scarantino, A. (2013). Rethinking functional reference. *Philos. Sci.* 80, 1006–1018. doi: 10.1086/673900
- Scheider, L., Waller, B. M., Oña, L., Burrows, A. M., and Liebal, K. (2016). Social use of facial expressions in hylobatids. *PLoS One* 11:e0151733. doi: 10.1371/journal.pone.0151733
- Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., and Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS One* 8:e76674. doi: 10.1371/journal.pone.0076674
- Schel, A. M., Tranquilli, S., and Zuberbühler, K. (2009). The alarm call system of two species of black-and-white colobus monkeys (*Colobus polykomos* and *Colobus guereza*). *J. Comp. Psychol.* 123, 136–150. doi: 10.1037/a0014280
- Scott-Phillips, T. C. (2008). “On the correct application of animal signalling theory to human communication,” in *Proceedings of the 7th International Conference (EVOLANG7): The Evolution of Language*, eds D. M. Smith, R. F. Smith, and R. F. Cancho (Singapore: World Scientific), 275–282. doi: 10.1142/9789812776129_0035
- Scott-Phillips, T. C. (2015a). Nonhuman primate communication, pragmatics, and the origins of language. *Curr. Anthropol.* 56, 56–80. doi: 10.1086/679674
- Scott-Phillips, T. C. (2015b). *Speaking Our Minds: Why Human Communication Is Different, and How Language Evolved to Make It Special*. New York: Palgrave MacMillan. doi: 10.1007/978-1-137-31273-0
- Scott-Phillips, T. C. (2016). Meaning in great ape communication: summarising the debate. *Anim. Cogn.* 19, 233–238. doi: 10.1007/s10071-015-0936-3
- Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 210, 801–803. doi: 10.1126/science.7433999
- Sievers, C., and Gruber, T. (2016). Reference in human and non-human primate communication: what does it take to refer? *Anim. Cogn.* 19, 759–768. doi: 10.1007/s10071-016-0974-5
- Slocombe, K. E., Kaller, T., Call, J., and Zuberbühler, K. (2010). Chimpanzees extract social information from agonistic screams. *PLoS One* 5:e11473. doi: 10.1371/journal.pone.0011473
- Slocombe, K. E., Waller, B. M., and Liebal, K. (2011). The language void: the need for multimodality in primate communication research. *Anim. Behav.* 81, 919–924. doi: 10.1016/j.anbehav.2011.02.002
- Smith, W. J. (1965). Message, meaning, and context in ethology. *Am. Nat.* 99, 405–409. doi: 10.1086/282382
- Tanner, J. E., and Byrne, R. (1996). Representation of action through iconic gesture in a captive lowland gorilla. *Curr. Anthropol.* 37, 162–173. doi: 10.1086/204484
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: The MIT Press.
- Tomasello, M., Call, J., Nagell, K., Olguin, R., and Carpenter, M. (1994). The learning and use of gestural signals by young chimpanzees: a trans-generational study. *Primates* 35, 137–154. doi: 10.1007/BF02382050
- Tomasello, M., Gust, D., and Frost, G. T. (1989). A longitudinal investigation of gestural communication in young chimpanzees. *Primates* 30, 35–50. doi: 10.1007/BF02381209
- Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., et al. (2017). Exorcising Grice's ghost: an empirical approach to studying intentional communication in animals. *Biol. Rev.* 92, 1427–1433. doi: 10.1111/brev.12289
- Vea, J. J., and Sabater-Pi, J. (1998). Spontaneous pointing behaviour in the wild pygmy chimpanzee (*Pan paniscus*). *Folia Primatol.* 69, 289–290. doi: 10.1159/000021640
- Waller, B. M., Caeiro, C. C., and Davila-Ross, M. (2015). Orangutans modify facial displays depending on recipient attention. *PeerJ* 3:e827. doi: 10.7717/peerj.827
- Wheeler, B. C., and Fischer, J. (2012). Functionally referential signals: a promising paradigm whose time has passed. *Evol. Anthropol.* 21, 195–205. doi: 10.1002/evan.21319
- Wilke, C., Kavanagh, E., Donnellan, E., Waller, B. M., Machanda, Z. P., and Slocombe, K. E. (2017). Production of and responses to unimodal and multimodal signals in wild chimpanzees, *Pan troglodytes schweinfurthii*. *Anim. Behav.* 123, 305–316. doi: 10.1016/j.anbehav.2016.10.024
- Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Anim. Behav.* 63, 293–299. doi: 10.1006/anbe.2001.1914

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Liebal and Oña. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Visuo-Kinetic Signs Are Inherently Metonymic: How Embodied Metonymy Motivates Forms, Functions, and Schematic Patterns in Gesture

Irene Mittelberg*

Natural Media Lab, Center for Sign Language and Gesture and Institute of English, American and Romance Studies, RWTH Aachen University, Aachen, Germany

OPEN ACCESS

Edited by:

Wendy Sandler,
University of Haifa, Israel

Reviewed by:

Seana Coulson,
University of California, San Diego,
United States
Antonio Barcelona,
Universidad de Córdoba, Spain
Pamela Permiss,
University of Brighton,
United Kingdom

*Correspondence:

Irene Mittelberg
mittelberg@humtec.rwth-aachen.de

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 01 January 2018

Accepted: 25 January 2019

Published: 27 February 2019

Citation:

Mittelberg I (2019) Visuo-Kinetic Signs Are Inherently Metonymic: How Embodied Metonymy Motivates Forms, Functions, and Schematic Patterns in Gesture. *Front. Psychol.* 10:254. doi: 10.3389/fpsyg.2019.00254

This paper aims to evidence the inherently metonymic nature of co-speech gestures. Arguing that motivation in gesture involves iconicity (similarity), indexicality (contiguity), and habit (conventionality) to varying degrees, it demonstrates how a set of metonymic principles may lend a certain systematicity to experientially grounded processes of gestural abstraction and enaction. Introducing *visuo-kinetic signs* as an umbrella term for co-speech gestures and signed languages, the paper shows how a frame-based approach to gesture may integrate different cognitive/functional linguistic and semiotic accounts of metonymy (e.g., experiential domains, frame metonymy, contiguity, and pragmatic inferencing). The guiding assumption is that gestures metonymically profile deeply embodied, routinized aspects of familiar scenes, that is, the motivating context of frames. The discussion shows how gestures may evoke frame structures exhibiting varying degrees of groundedness, complexity, and schematicity: basic physical action and object frames; more complex frames; and highly abstract, complex frame structures. It thereby provides gestural evidence for the idea that metonymy is more basic and more directly experientially grounded than metaphor and thus often feeds into correlated metaphoric processes. Furthermore, the paper offers some initial insights into how metonymy also seems to induce the emergence of schematic patterns in gesture which may result from action-based and discourse-driven processes of habituation and conventionalization. It exemplifies how these forces may engender grammaticalization of a basic physical action into a gestural marker that shows strong metonymic form reduction, decreased transitivity, and interacting pragmatic functions. Finally, addressing basic metonymic operations in signed lexemes elucidates certain similarities regarding sign constitution in gesture and sign. English and German multimodal discourse data as well as German Sign Language (DGS) are drawn upon to illustrate the theoretical points of the paper. Overall, this paper presents a unified account of metonymy's role in underpinning forms, functions, and patterns in visuo-kinetic signs.

Keywords: gesture, metonymy, frames, scenes, iconicity, contiguity, indexicality, schematicity

INTRODUCTION

Gestures are essentially metonymic: Iconic gestural figurations and enactments, in particular, exhibit the principle of partial semiotic portrayal par excellence. In interaction with concurrent speech, evanescent hand shapes and movements tend to abstract salient characteristics from, briefly allude to, or otherwise evoke entire persons, three-dimensional objects, holistic motion events, and rich contexts (e.g., Gibbs, 1994; Müller, 1998; Bouvet, 2001; Mittelberg, 2006; Calbris, 2011). With their gestures and postures, speakers typically foreground the particular aspects of previously witnessed or newly imagined objects, actions, behaviors, or scenarios that are especially relevant to their communicative intentions in ongoing discourses. They may trace, for instance, the spatial proportions of a building in the air or imitate a person's action, such as running to catch a bus or handing a present to someone, in a reduced or stylized fashion. Interlocutors may thus, consciously or not, convey essential facets and kinesthetic qualities of their embodied experiences, memories, habits, mental imagery, or the immediate environment by schematically but effectively making them tangible and thus intersubjectively sharable in the here and now of a given multimodally orchestrated speech event (e.g., Sweetser, 2007; Mittelberg, 2013; Müller, 2014, 2017).

For example, if I am telling a friend that I will be spending the entire weekend working on my paper and simultaneously make a fleeting typing action, my hands simulate typing on an imaginary keyboard. From such a quick iconic gestural action, the addressee may infer that I will, in fact, be carefully and concentratedly typing for hours on the keyboard that actually exists on my desk. She can also infer the fact that, in this context, "working" means writing with the help of a computer. Moreover, she can imagine the written text that will result from this action, the content of the paper she knows I am working on, as well as other practically and ideationally related actions, entities, stages, versions, and mental or emotional states involved in eventually reaching the goal of submitting the finalized manuscript. All these various aspects are metonymically linked in a pragmatically structured context of experience, or *frame* (Fillmore, 1982), in which one gesture may evoke not only the immediately contiguous virtual keyboard, but also trigger an ensuing associative chain and a larger semantic network (e.g., Calbris, 2011; Mittelberg and Waugh, 2014; Mittelberg, 2017a).

Metonymic Motivation of Gestural Abstraction and Enaction: More Than Iconicity

The primary aim of this paper is to pinpoint the inherently metonymic nature of co-speech gestures. It will show how distinct metonymic principles may lend a certain experientially grounded regularity to processes of *ad hoc* abstraction and enaction that are involved in gestural sign formation (e.g., Arnheim, 1969; Müller, 1998). Gestural abstraction and the resulting schematicity here are assumed not to be random, but experientially, cognitively, linguistically, pragmatically, and culturally motivated. Due to the temporal dynamics of face-to-face communication, there is only

a very limited amount of time to perform gestures in sync with the conceptual contents, incremental articulation, and prosodic contours of the simultaneously evolving utterance as well as with other bodily signs such as gaze and head movements.

Crucially, visual perception, often somewhat privileged in cognitive approaches to language, is only one sensory, experiential source from which gesturers intuitively draw their semiotic material. A point the present proposal wishes to make is that motivation in gesture involves more than iconicity (e.g., McNeill, 1992; Mittelberg, 2006, 2014; Perniss et al., 2010).¹ It is claimed that in gesture, besides interacting with iconicity and metaphoricality, it is through indexicality that metonymy also operates on latent contiguity relations between the hands and the material and social world.² Such contiguity relations may become operationalized for gestural communication and thus lead the interpreting mind to 'grasp' the virtual objects and tools that gesturing hands seemingly manipulate (e.g., Müller, 1998; Streeck, 2009; Mittelberg and Waugh, 2014). By laying out how gestures may metonymically evoke frames through picking out aspects of *basic scenes of experience* (Fillmore, 1977, 1982; Goldberg, 1995, 1998), it will further be argued that a frame-based approach to gesture may not only integrate various accounts of metonymy (e.g., Dancygier and Sweetser, 2014), but also account for processes of pragmatic inferencing that are often heavily involved in gesture interpretation.

While the paper focuses on spontaneous gestures that are temporally, semantically, and syntactically integrated with concurrent speech (e.g., McNeill, 1992; Müller, 1998; Kendon, 2004; Fricke, 2012), metonymic modes underpinning iconic signs in sign language will also be addressed to highlight some commonalities regarding principles of sign constitution. Furthermore, the paper offers some initial insights into how embodied metonymic principles also seem to underpin discourse-pragmatic processes of routinization and schematization in gesture; that is, how metonymy may induce the emergence of gestural patterns with increased degrees of habit-driven conventionalization (e.g., Mittelberg, 2006, 2017a,c; Müller, 2017). Overall, this paper presents the first unified account of metonymy's role in underpinning forms, functions, and patterns in visuo-kinetic signs.

Gestures as Visuo-Kinetic Signs in Multimodal Contextures of Communicative Action

Co-speech gestures here are understood as discourse-embedded, kinetic action that is performed with the head, hands, arms, torso, or entire body and has some communicative function(s)

¹According to Peirce (1960, p. 157), "icons have qualities which resemble those of the objects they represent, and they excite analogous sensations in the mind." The term 'Object' here encompasses existing and imagined entities, persons, actions, etc. For both the gesturer and the interpreter, gestural icons thus rely on a *perceived* similarity between the gestural form and what they are signify.

²In indexical signs, the relation between sign and object is based on contiguity, according to Peirce, that is, on a factual (physical or causal) connection between a gestural form and what it evokes: "An index is a sign that denotes its object by virtue of being really affected by that object" (Peirce, 1960, p. 143). See also Table 1.

(e.g., Müller, 1998; Kendon, 2004; Calbris, 2011). Partly in reference to Jakobson (1972, p. 474) notion of “motor signs,” the term *visuo-kinetic signs* is introduced here to encapsulate the fact that gestures are part of, or emerge from, the human body with its inherent morphology, motion range, motor routines, and multiple senses with which we perceive and understand the world around us. Gestures genuinely preserve and (re-)enact some of their kinetic, sensorimotor, tactile, and interpersonal origins (e.g., Mittelberg, 2010, 2018; Müller, 2017). While gestures usually need concurrent speech to specify their local meaning, they often *do* something in their own right and in their own specific, experientially motivated ways (e.g., Mittelberg and Joue, 2017; Müller, 2017; Wehling, 2017).

While gestures are part of the visual – and thus visible and observable – facets that make up contextualized language use in interaction, the Kendon’s (2004) idea of *visual action as utterance* duly emphasizes the fact that gestures are more than just visual. Gestures are communicative *bodily actions* that are instantaneously *performed* by human beings and dynamically *evolve* in time and space (e.g., Müller, 1998; Sweetser, 2007; Goodwin, 2011; Streeck et al., 2011). One important factor in gesture interpretation and analysis, however, resides in the fact that the ‘semiotic material’ we are looking at consists not only of observable physical components – such as body posture, body motion, finger configurations, as well as the position and movements of gesturing hands – but also of immaterial, yet signifying, components such as evanescent movement traces created in the air or imaginary surfaces, objects, or points in space (e.g., Mittelberg, 2010; Hassemer, 2015). Speakers’ hands often pretend to hold or otherwise manipulate virtual objects and/or tools – the typing gesture necessarily implies an imagined keyboard – or to interact with imaginary interlocutors. Consequently, to do justice to the noted specific semiotic nature of gestures, the present account of gestures as visuo-kinetic signs also includes elements and dimensions of multimodally achieved sign processes that are not visual, and hence rather invisible, but still contribute to a gesture’s kinesthetic feel, meaning, and pragmatic function(s) (e.g., Mittelberg, 2006, 2013). As will be shown below, metonymy enables us to account for the virtual elements thus implied, or created on-the-fly, which may be inferred from their dynamically evolving multimodal *semiotic contextures* (Jakobson, 1956; see also Müller, 1998; Streeck, 2009; Goodwin, 2011). By illuminating the pragmatic workings of metonymy in visuo-kinetic signs, this paper seeks to provide additional insights into the nature of both gesture and metonymy.

TOWARD A FRAME-BASED ACCOUNT OF EMBODIED METONYMY IN GESTURE

Metonymy belongs – together with metaphor, synecdoche, and irony – to the four master tropes (Burke, 1941). Jakobson (1956)

was one of the first to advocate a balanced theory of metaphor and metonymy as two universal principles of association and signification that are prominent in language, thought, discourse, literature, and the visual arts (e.g., Waugh and Monville-Burston, 1990). Subsequently, experientialist views on language and the embodied mind attributed a preeminent role to metaphor (e.g., Lakoff and Johnson, 1980; Johnson, 1987; Lakoff, 1987; Sweetser, 1990). The ground-laying idea was that the human conceptual system, language, language change, and language use, encompassing all types of discourse, are structured and function metaphorically to an extent that had previously been underestimated.

With little delay, metonymy has become recognized as an equally important figure of thought and language (e.g., Gibbs, 1994, 1999; Panther and Radden, 1999; Barcelona, 2000a,b, 2003; Dirven and Pörings, 2002; Fauconnier and Turner, 2002; Panther and Thornburg, 2003). In recent years, cross-linguistic research has clearly confirmed that metonymy plays a constitutive role in conceptual, semantic, and grammatical structuring, as well as in discourse processes, including, for example, indirect reference, speech acts, and pragmatic inferencing (e.g., Barcelona, 2009; Panther et al., 2009; Benczes et al., 2011; Kövecses, 2013; Littlemore, 2015). A crucial tenet of the present proposal on how gesturally engendered sign processes involve ‘metonymy in the making’ is that “metonymy is a central organizing principle of pragmatics, the contextual use and interpretation of meaning” (Dancygier and Sweetser, 2014, p. 162; see also contributions in Hampe, 2017).

Furthermore, there is a growing body of work on metonymy in various modalities, media, and art forms, ranging from mnemonic devices, painting, material culture to advertisement and film (e.g., Jakobson and Pomorska, 1983; Mittelberg, 2002, 2006; Forceville, 2007; Forceville and Urios-Aparisi, 2009; Littlemore, 2015). Regardless of the modality or medium in which metonymy materializes, it may create single meaningful sparks in our minds or set into motion complex associative chains and networks (e.g., Benczes et al., 2011). Metonymy thus may propel diverse processes of reasoning, imagination, and discourse construction (e.g., Coulson, 2001; Fauconnier and Turner, 2002; Dancygier and Sweetser, 2014), both within one modality and across modalities.

Due to limits of space, this paper cannot provide a comprehensive overview of all the different kinds, functions, and manifestations of metonymy described in the literature. Rather, I will draw on the approaches to metonymy that seem particularly apt to account for the structuring and meaning-making processes at work in bodily signs that partake in multimodal interaction. I will thus try to show why, in the case of manual gestures and other visuo-kinetic signs, it makes sense to shift the focus from strongly cognitively oriented accounts to truly embodied, or body-based, understandings of metonymy. To this effect, the exposition below will provide further gestural evidence for the claim that metonymy is more directly experientially motivated than metaphoric processes with which they tend to interact (building on Mittelberg, 2006, 2013, 2017a; Mittelberg and Waugh, 2009, 2014; see also, e.g., Barcelona, 2000b; Kövecses, 2013; Dancygier and

Sweetser, 2014; Ruiz de Mendoza Ibanez, 2017).³ Advocating a frame-based account of metonymy in gesture, the ensuing sections aim to show how various metonymic principles function as fundamental construal mechanisms that drive pragmatically grounded processes of embodied schematization in co-speech gestures.

Experiential and Functional Domains

According to domain-based accounts, metonymic mappings occur within the same cognitive or experiential domain, or within the same *idealized cognitive model* (i.e., ICM, Lakoff, 1987; Panther and Radden, 1999: 19ff.). Barcelona (2003, p. 83) provides the following definition with a functional emphasis: “(a) metonymy is a mapping of a cognitive domain, the source, onto another domain, the target. Target and source are in the same functional domain and are linked by a pragmatic function, so that the target is mentally activated.” Let us consider the by now classic example of a metonymic linguistic expression given in (1) (Lakoff and Johnson, 1980, p. 35; see also Dancygier and Sweetser, 2014, p. 5):

- (1) *The ham sandwich is waiting for his check.*
- (2) *Table 5 urgently needs to pay.*

Here, “the ham sandwich” does not refer to a food item but, indirectly, to the restaurant client who ordered it. The dish previously served by a member of the service personnel, and probably already consumed by the client, thus stands for the latter, based on contextual, pragmatic links binding these elements within one and the same functional domain (e.g., Fauconnier and Turner, 2002). Another common way to refer to restaurant clients is by the number of the table they are sitting at, such as in (2). Here, a different element in this particular experiential domain or pragmatic context is highlighted, namely the table as a physical location inside the restaurant. In both expressions, the client is the metonymic target domain. The kind of domain that is chosen to be the metonymic source domain depends on the pragmatic forces and customs at work in a particular context of use. The factors that determine the choice of source domain in (1) and (2) include the interpersonally built-up common ground and the professional practices of the service personnel, who are used to communicating about this kind of frequently occurring situation.

As is well known, metaphor, by contrast, involves a mapping between two different experiential domains, as expressed by, for instance, the conceptual metaphor UNDERSTANDING IS GRASPING (e.g., Lakoff and Johnson, 1980). This cross-domain mapping gives rise to metaphoric linguistic expressions, such as in (3), where the abstract target domain of understanding

is conceptualized in terms of the physical source domain of manually seizing an object. Alternatively, the same target domain may be structured by another bodily source domain, that is, visual perception as a way of comprehending something (UNDERSTANDING IS SEEING, *ibid.*), such as in (4).

- (3) *Paula grasped the new idea right away.*
- (4) *Paula instantly saw what I meant.*

In gesture, body-centered and action-based source domains may intuitively activate pragmatic links to metonymic targets with which they are connected through repeated instances of similar physical experience (e.g., Mittelberg and Joue, 2017). Certain manual actions may evoke the objects or tools that are routinely handled when they are actually performed. For instance, to ask for more bread in a restaurant, one may first raise a hand to catch the waiter’s attention and then point with that hand at the empty bread basket one is holding up with the other hand. The waiter will readily understand this gestural request based on the gesture and the empty bread basket, which here functions as the source domain pointing to the desired target: additional bread. Put differently, the CONTAINER stands metonymically for the wanted CONTENT; arriving at the latter involves following a contextually shaped, inferential pathway (e.g., Barcelona, 2003, 2009; Panther and Thornburg, 2003). Bread basket and bread belong to the same mundane experiential domain not only in people’s homes, but also in a restaurant context, where it is common practice to serve bread in baskets and also to provide refills.

If one performs this gestural request without concurrent speech, the basket metonymically evokes the idea of bread on visual and experiential grounds. However, it is likely that the person wanting more bread actually also verbally asks the waiter for it, as in (5) or (6), once the latter has arrived at the table. In both linguistic examples, there is no mention of the basket, only of the (non-existent) bread. Furthermore, (6) functions as an indirect speech act, that is, an assertion indirectly functions as a request in this case. Qualifying as a *speech act metonymy*, it may be understood “as a *scenario* having metonymic structure” (Panther and Thornburg, 2003, p. 128).

- (5) *Could we have more bread please?*
- (6) *We are out of bread.*

Building on Langacker (1987), Croft (1993) extended the single-domain approach to a domain matrix, which involves shifts in foregrounding from one domain to another domain in the same matrix. Applying this idea to the bread-request scenario, we can assume that first the empty basket is foregrounded due to its perceptual prominence with respect to the bread that formerly existed in it; then, the metonymic process causes a shift in foregrounding onto the metonymic target, that is, the indirectly referenced bread that the client would like the waiter to fetch.

When it comes to gesture and multimodal interaction, the focus is naturally on the communicating human body and thus especially on those experiential domains that are indexically anchored in the material and social contexts of people engaged

³The line of research on which this paper builds (Mittelberg, 2006, 2008, 2013, 2014, 2017a; Mittelberg and Waugh, 2009, 2014) combines traditional semiotic frameworks that are not exclusively based on language – notably the works of Jakobson (1956, 1972) and Peirce (1960) – with embodied approaches to language, cognition, and interaction (e.g., Fillmore, 1982; Johnson, 1987; Lakoff, 1987; Goldberg, 1995; Gibbs, 2006; Dancygier and Sweetser, 2014). While the previous work focused on semantic and pragmatic functions of metonymy in gesture, this paper is the first account that integrates related aspects of embodied grammaticalization, multimodal constructions, and lexemes in signed languages.

in some sort of physical action or in communicative exchange (Mittelberg, 2017a). For example, a participant in a study on transitive action gestures (Grandhi et al., 2011) gives the following verbal instruction regarding how to slice an apple into pieces:

(7) *You need to slice the apple by holding it down and cutting it there.*

Here no actual (i.e., visible) physical object is involved in the gestural portrayal. Pretending to be holding a virtual knife in her dominant right hand, she pantomimes how she would cut a virtual apple that she is seemingly holding down with her left hand (Figure 1, adapted from Grandhi et al., 2011).⁴

Indeed, slicing an apple into pieces necessitates a particular action (cutting), an object (apple), and a tool (knife). All three elements thus belong to the same experiential domain or scenario. Whereas what the participant says in (7) draws attention to the cutting action and the object, but not to the tool, the latter can be easily inferred from the action context. Again, the actions and objects implied belong to an everyday domain of experience. Moreover, this example involves a CAUSE-EFFECT metonymy, for we can imagine the apple first in its entirety and then the slices resulting from the cutting action (as shown in Figure 1; see also Mittelberg and Waugh, 2014).

Semantic Frames and Familiar Scenes of Experience

A large part of what has been described above based on cognitive, experiential, or functional domains can also be understood in frame-semantic terms (see also, e.g., Panther and Radden, 1999, p. 9; Kövecses, 2013; Dancygier and Sweetser, 2014). According to Fillmore (1982, p. 111), the term *frame* covers “any system of concepts related in such a way that to understand any of them you have to understand the whole structure in which it fits.” Frames can thus be understood as metonymically structured wholes in which one of its parts may evoke another correlated part or the frame as a whole (e.g., Mittelberg and Joue, 2017).

⁴ An ethics approval was not required as per applicable guidelines and regulations. Written informed consent was obtained from all participants. For all figures appearing in this paper, written informed consent was obtained from the participants for publication of the images.

While the semantic structures in question are situated at relatively high levels of abstractness, Fillmore (1975, p. 127) emphasizes how frames are experientially grounded in familiar *scenes* which underpin the acquisition of word meanings and the gradual differentiation of whole scenarios into their constitutive parts. Scenes “include not only visual scenes but familiar kinds of interpersonal transactions, standard scenarios, familiar layouts, institutional structures, enactive experiences, body image; and, in general, any kind of coherent segment, large or small, of human beliefs, actions, experiences, or imaginings” (Fillmore, 1977, p. 63). Since human behavior and gestures are intrinsic to such scenes and are also shaped by them, it seems fitting to exploit the notions of both frames and scenes to explicate gestural communication (e.g., Sweetser, 2012; Mittelberg and Waugh, 2014). As proposed in earlier stages of the present account (e.g., Mittelberg, 2017a,c; Mittelberg and Joue, 2017), gestures that recruit frame structures tend to metonymically pick out essential elements and salient qualities of scenes, that is, the motivating context of frames. This especially pertains to situated factors of real-world, enactive experiences that can be recruited for both literal and metaphorical construal and thus also involve *primary scenes* and *primary metaphor* (Grady, 1997; see also section “Metonymy Underpins Schematic Gestural Patterns and Fully Codified Visuo-Kinetic Signs”). The ways in which embodied metonymy plays a central role in frame-based processes that drive multimodal discourse pragmatics is discussed in the next section.

Frames and Frame Metonymy in Co-speech Gestures

Dancygier and Sweetser (2014: 102ff.) point out that, compared to domains, the structural organization of frames allows for a more fine-grained and systematic account of correlations not only within a frame (thus giving rise to *frame metonymy*), but also between two frames that are partially mapped onto each other (thus giving rise to metaphor). They provide the following general definition of metonymy: “the use of some entity A to stand for another entity B with which A is *correlated*” (ibid.: 134, italics in the original). Frame metonymy refers “to all usages where one reference to an element of a frame is used to refer to either the frame as a whole or to other associated elements of the frame” (ibid.: 135), for example, where ‘the Crown’ refers to the British monarchy. *Part-whole* frame metonymy includes what

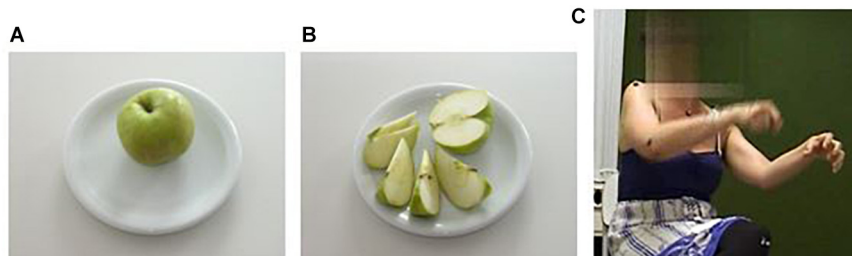


FIGURE 1 | Photos (A,B) show a before-and-after scenario presented to participants who were asked to tell interviewers how to get from (A) to (B). The video still (C) shows how gesture draws on the experiential domain of cutting an apple into slices; adapted from Grandhi et al. (2011). Written informed consent was obtained from the depicted individuals for the publication of this image.

is generally understood by *synecdoche*, for example, where ‘field hands’ stands for people who work outdoors on a farm.

An often-cited example is the RESTAURANT DINING frame (or *script*, see Schank and Abelson, 1977); it implies elaborated scenarios involving certain culturally shaped sets of elements, roles, behaviors, and sequences of events (Fillmore, 1982). Seen from this perspective, Examples (1), (2), (5), and (6) discussed in Section “Experiential and Functional Domains” involve items that are integral to this frame structure: the guest who ordered the ham sandwich, the sandwich itself, Table 5, the bread basket, ordering more bread, and asking for the check. We are able to place and relate all these items within a structured, dynamic fabric of correlations that allows us to quickly understand acts of indirect reference and other metonymic operations occurring within it. Such larger frame structures, or scripts, are supposed to be active in the background processing of cognition and behavior (e.g., Coulson, 2001), in the sense that one becomes aware of them if an element is omitted or occurs sequentially out of place; for example, if someone asks for the bill before having consumed the dish that s/he ordered.

In processes of frame-based language use, reasoning, and discourse understanding, networks of metonymic relations inherent to specific frames thus become activated and operationalized (e.g., Coulson, 2001; Dancygier and Sweetser, 2014). Thereby, each frame structure provides various springboards for metonymic associations as well as entry points and conceptual bridges for intersubjective meaning construction in conversational exchanges or collaborative story telling. In ongoing interaction, speakers may use linguistic, gestural, or eye-gaze cues to frame a given scene in positive, critical, doubtful, or humorous terms, from a scene-internal or a scene-external viewpoint, or by adopting several viewpoints simultaneously (e.g., McNeill, 1992; Dudis, 2004; Sweetser, 2012; see Mittelberg, 2017b on the interplay of viewpoint, indexicality, and metonymy in gesture). Alluding to a particular discourse-relevant frame element may automatically trigger connections of different scope and varying complexity, for example, to directly correlated elements, the frame as a whole, or metaphoric associations.

With regard to the typing gesture described in the introductory section, a decisive detail lies in the fact that understanding the message involves a *cross-modally* instantiated, frame-internal metonymic process. So, again, if I imitate typing with both hands while saying to a friend:

(8) *I’ll be working the entire weekend on my paper.*

the pantomimed action of typing not only gets profiled against the ground of the imaginary keyboard, thus evoking the TYPING and WRITING frames, but also against the backdrop of larger frame structures, such as WRITING AN ACADEMIC PAPER or PUBLISHING. Note that in the verbal part of the utterance, the verb does not refer to the gesturally simulated typing action, but to the more general WORK frame. Hence, a cross-modal metonymic process takes place whereby the gesture specifies the verbally communicated information ‘I’ll be working’ as ‘typing’ or ‘writing’ a manuscript. For the interlocutors, this visuo-kinetic sign (including the imaginary keyboard) may instantly serve as a

dynamically created *material anchor* (Hutchins, 2005) for joint attention and thus evoke aspects of their shared experience of such situations. In this way, webs of associations may branch out from such a mutual gestural anchor: In their respective embodied minds, this may facilitate associations that are not only directly grounded in physical experience, such as manipulating a keyboard or touchpad, but also bring to mind less tangible associations, such as subsequent phases of the work process, the potential structure and content of the paper, a previously co-authored paper, as well the community’s reaction (see also Calbris, 2011: 10f.). Through activating the WEEKEND frame, they might also think of what one misses out on while working the entire weekend. For both the speaker and the interlocutor(s), frame-metonymic associations are thus also likely to solicit subjective and intersubjective dimensions connected with certain mental or emotional states, such as being focused, anxious, or happily working away (see also Mittelberg and Waugh, 2014).

Regarding linguistic expressions, Dancygier and Sweetser (2014, p. 108) further emphasize that a certain degree of salience is needed to clearly associate a term with a frame in the sense of Langacker’s (1987) notion of *active zone* as the profiled part of a whole.⁵ For a body-based and action-based view of metonymic processes (Mittelberg, 2017a), it is particularly relevant that the human body forms a metonymically structured whole in and of itself. Certain parts of it, for example, the head or hands, may become prominent in a meaning-making process such as in the verbal example of ‘field hands’ mentioned above. In this kind of part-for-whole frame metonymy “the part centrally or directly involved in an activity stands for the whole. The hand, for example, is the part of the arm used for holding, touching, etc.; hence it is the active zone of the arm for many purposes” (Dancygier and Sweetser, 2014, p. 144).

Visuo-kinetic signs performed by heads, shoulders, and/or hands may also function as the signifying, active zone of the gesturer’s body that ‘stands out’ within dynamic multimodal contextures and may thus become meaningful. Furthermore, these signs may metonymically stand for the entire person making the gesture – or for the belief system behind a certain stance s/he is expressing toward what is being said (see also Calbris, 1990 on *body segments*).⁶

Gestural Frame Evocation at Varying Levels of Groundedness and Complexity

Building on Fillmore’s (1977, 1982) notion of semantic frames, Mittelberg (2017a) has recently presented a frame-based account of gesture pragmatics. It proposes different kinds of embodied frame structures that are situated at varying levels of groundedness, schematicity, and complexity, a synopsis of which will be presented here.

First, *basic physical action frames* and *basic object frames* are understood as being directly grounded in physical experience and

⁵Langacker (2009, p. 48) defines *active zones* as follows: “An entity’s active zone, with respect to a profiled relationship, is that facet of it which most directly and crucially participates in that relationship.”

⁶In sign language, *body partitioning* (Dudis, 2004) also draws on the metonymic organization and affordances of the signer’s body to adopt and combine different viewpoints on a given scene.

basic scenes (Mittelberg, 2017a: 215ff.). These strongly embodied frames mainly encompass *prototypical events* (Slobin, 1985) such as pushing, pulling, and teasing apart; *mimetic schemas* (Zlatev, 2014) such as jump, kick, grasp, and hit; *basic-level actions* (Lakoff, 1987) such as eating, running, and walking; as well as any other intransitive, transitive or ditransitive actions that may be simulated via gestures and whole-body enactments (e.g., Hostetter and Alibali, 2008; Bressem and Müller, 2014; Müller, 2017). In addition, basic physical action frames may intertwine with basic object frames to account for the physical entities that the former, together with their affordances, typically imply (as in **Figures 1, 4**; see also Grady, 1997 on primary scenes).⁷ Basic object frames also get evoked in multimodal descriptions of physical entities or spaces.

Second, more *complex frame structures* comprise frames that are internally more differentiated, more detached from the motivating contexts of experience and hence situated at higher levels of abstractness (Mittelberg, 2017a: 220ff.). Presupposing the “development of a complex frame out of correlated simpler frames” (Dancygier and Sweetser, 2014, p. 138), we will first consider frame structures that are composed of several connected basic action or object frames and hence exhibit an intermediate level of groundedness. The RESTAURANT DINING script (see section “Frames and Frame Metonymy in Co-speech Gestures”), for example, consists of such a culturally shaped ordered set of basic actions and their implied objects and/or interacting persons that are fairly well grounded and may thus function as experiential anchors: being seated, looking at the menu, signaling to the waiter, eating, paying, etc. Each of the sequenced actions and behaviors involve physical activities and can thus be easily enacted through postures, gestures, and facial mimics, and hence evoke other, correlated items or the overarching frame as a whole.

Highly abstract complex frame structures are understood as being a lot more detached from motivating contexts than the frames discussed so far. They involve cognitive and semiotic structures and activities that people rely on when producing or describing phenomena at a meta-level, for instance, linguistic structures, genre-dependent narrative and conversational patterns, plots of novels, films, or animated cartoons, mental maps, as well as knowledge systems and schematic conceptual structures such as theories or category systems (Mittelberg, 2017a: 223f.). In gesture, such larger architectures may be overtly represented and thus become visible, albeit minimally and fleetingly, via virtual time lines traced in the air (e.g., Calbris, 2011) or other diagrammatic configurations of points placed in gesture space that highlight how individual words, items, places, events, concepts, or more general discourse contents relate to one another temporally, spatially, or logically (e.g., Kendon, 2004; Mittelberg, 2008; Enfield, 2009; Bressem, 2014). Beat gestures (McNeill, 1992) are also a means to accentuate particularly relevant parts of an utterance, thus making them metonymically stand out from the speech chain as a whole.



FIGURE 2 | Speaker evoking the SWIMMING action frame in a reduced and partial metonymic fashion. Written informed consent was obtained from the depicted individuals for the publication of this image.

Let us now look at how basic and more complex frames may interact in organizing a thematic unit of multimodal discourse. Example (9) is taken from a description of a past vacation scene produced in the context of a travel-planning task. Suggesting Hungary as a possible destination on a joint trip through Europe, the person on the right in **Figure 2** is telling her conversational partner in German that on a previous visit to Budapest the weather was very nice. By mentioning that it was beautiful outside (‘it... was really nice’), the speaker verbally evokes the WEATHER frame in an indirect fashion. She then profiles a specific sub-frame against the backdrop of the larger, general WEATHER frame by uttering the compound noun ‘*Schwimmwetter*’ (swimming weather).⁸

- (9) *es (...) war richtig schön und so [Schwimmwetter]*
(it (...) was really nice and like [swimming weather])

The simultaneously produced visuo-kinetic sign shown in **Figure 2**, consisting of simulated swimming movements performed by the speaker’s hands and arms, renders this specification salient. This gesture can be said to activate the *basic physical action frame* SWIMMING. Not all the body parts usually involved in swimming participate in this partial, stylized iconic enactment of a learned motor routine. With reference to Hostetter and Alibali’s (2008) *gestures-as-simulated-action* (GSA) framework, this gestural action exemplifies how “gestures emerge from the perceptual and motor simulations that underlie embodied language and mental imagery” (Hostetter and Alibali, 2008, p. 502). Under the present view, this is another example of how communicative gestures may metonymically evoke the physical actions they are imitating iconically by only minimally enacting the onset or some essential characteristics of a full-blown action routine in a rather schematic fashion.

Looking at the immediate discourse context of this bimodal performance reveals that it belongs to a vivid description, provided in (10), in which several gestures portray additional aspects that belong to what seems a general, yet culture-dependent, understanding of WARM WEATHER. Enquiring

⁷See Mittelberg and Joue (2017) on gestural *source actions* as metonymic bases of metaphoric processes.

⁸Another way to express such a relation is to say that a detail is profiled against the backdrop of a cultural model (Cienki, 1998) or an idealized cognitive model (Lakoff, 1987).

about the weather conditions on this past trip, the participant on the left actually first evokes the WINTER frame: He asks whether there was snow and simultaneously makes a bimanual Palm-Up Open Hand gesture (PUOH, Müller, 2004), shown in **Figure 3A**. This pragmatic gesture functions here as a visuo-kinetic question marker, or an interactive *seeking* gesture (Bavelas et al., 1995), that is soliciting an answer from his interlocutor in an ‘empty-handed’ manner (see also Kendon, 2004; Streeck, 2009; Bressem and Müller, 2014). His interlocutor then replies that it was actually rather nice and warm.

- | | | | |
|------|----------------------|--|---------------------------------|
| (10) | P _{Left} : | lag Schnee? | [was there snow?] |
| | P _{Right} : | nein (–) es war schon warm | no (–) it was actually warm |
| | | also so | well like [t-shirt weather] and |
| | | [T-Shirt-Wetter] und [Shorts-Wetter] (...) | [shorts weather] (...) |
| | | war richtig schön | was really nice |
| | | und so | and like [swimming weather] |
| | | [Schwimmwetter]. | |

When explaining that it was “t-shirt weather” the speaker on the right rotates both hands at approximately shoulder height with the palms facing toward the t-shirt she is wearing (**Figure 3B**). This gesture may be interpreted as pointing to the short sleeves of her t-shirt. Considered as a *cyclic gesture* (Ladewig, 2011), it may also evoke the idea of continuously

feeling hot or of sensing hot air surrounding the body. The speaker then accompanies her verbal utterance “shorts weather” with another bimanual gesture: With the palms facing the torso, the outer edges of the hands indicate the location on her thighs where shorts typically end (**Figure 3C**). Only then does she multimodally activate the SWIMMING frame as described earlier (**Figures 2, 3D**). These individual, metonymically linked frame elements jointly draw on the WARM WEATHER frame as a whole. In this way, the semantic structures evoked in Example (10) involve several, metonymically correlated frame elements and are thus more complex than the individual basic physical action frame (SWIMMING) and the basic physical object frames (T-SHIRT and SHORTS) which constitute them. Larger frames at this intermediate level of groundedness are still rooted in habitual, mundane physical and social activities and thus may draw on various “scenes basic to human experience” (Goldberg, 1995, p. 5).⁹

Indeed, scenes have been found to be particularly relevant with respect to how interlocutors construe and follow processes of online meaning construction. According to Fillmore (1977, p. 1226), “in most natural conversations, the participants have, already ‘activated,’ a number of shared, presupposed, scenes that we can speak of as being in their consciousness as they speak.” This supports the idea that scenes partake in the dynamic

⁹For examples of highly abstract complex frame structures, see Mittelberg (2017a: 223ff.).

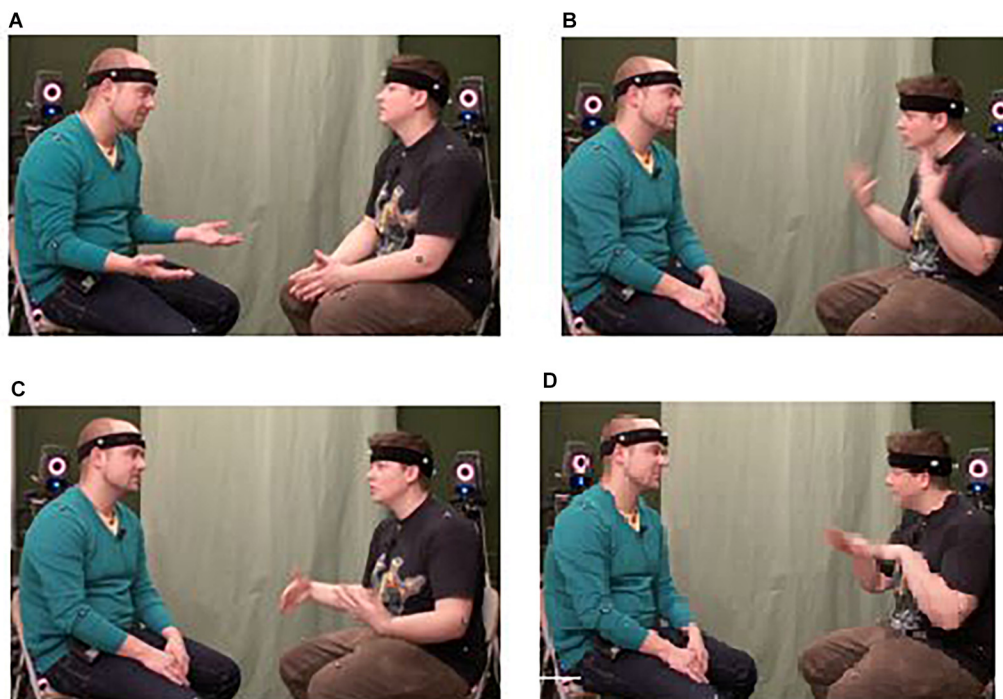


FIGURE 3 | Multimodal evocation of larger frame structures: **(A)** The linguistic question evokes the WINTER frame, while the gesture pragmatically evokes the QUESTION (speech act) frame; the linguistic reply evokes the WARM WEATHER frame characterized by a series of gestures, each of which details a different basic object/action frame: **(B)** T-SHIRT; **(C)** SHORTS, and **(D)** SWIMMING. Written informed consent was obtained from the depicted individuals for the publication of this image.

contexts that shape multimodal processes of conceptualization during both the production and interpretation of co-speech gestures. A frame-based understanding of multimodal discourse pragmatics has the advantage of including larger semantic networks that go beyond local reference or individual simulated actions, thus leading into discourse-driven processes of more complex meaning construction.

Although the different kinds of frame structures discussed so far only pertain to concrete actions and objects, they may, in principle, also underpin metaphoric construal in gesture (e.g., Dancygier and Sweetser, 2014; Mittelberg and Joue, 2017). This line of inquiry also leads into related linguistic issues such as how embodied scenes and frame metonymy may factor into related syntactic frames, grammaticalization in gesture, as well as multimodally instantiated constructions (e.g., Mittelberg, 2017c; Zima and Bergs, 2017; see section “Metonymy Underpins Schematic Gestural Patterns and Fully Codified Visuo-Kinetic Signs”).

Reference and Pragmatic Inferencing in Gesture

Exploring how metonymy motivates gestural practices of frame evocation necessarily raises questions concerning reference and inference. While these complex issues cannot be resolved here, let us pursue the idea that many gestures tend to *evoke frames* and *enact* or *simulate* physical actions rather than *represent* or *refer* to things or actions in the real world (e.g., Merleau-Ponty, 1962; McNeill, 2005; Mittelberg, 2018). Unlike spoken and signed languages, most spontaneous gestures do not rely on fully coded form-meaning pairings on the basis of which referential processes typically function. Rather, *habituated inferences* based on habitual actions as well as habits of gesture production and interpretation seem to play a central role in how gestures signify. Here, a parallel may be drawn with how metonymy plays a role in catalyzing inferential and referential interactions in language, as Barcelona points out:

Metonymy has this inferential role because of its ability to mentally activate the implicit pre-existing connection of a certain element of knowledge or experience to another. The referential function of metonymies is thus a useful (hence extremely frequent) consequence of their inference-guiding role since what we do when we understand a referential metonymy is to infer the referential intentions of others (Nerlich and Clarke, 2001; Barcelona, 2009, p. 369).

Metonymic inferences in co-speech gestures may occur within the gestural modality or cross-modally, that is, triggered by a linguistic cue (e.g., Mittelberg, 2017a). Experientially entrenched pragmatic inferences (Dancygier and Sweetser, 2014, p. 144) are indeed key to mentally simulating and understanding the communicative intentions of the gestural actions performed by others. As we saw earlier, **Figure 1** demonstrates a habitual metonymic correlation between gesturing hands and the cutting action they are simulating: the apple is seemingly being held down by one hand, while the seemingly held knife in the other is not referred to in the speech chain, but implied in the action. Performing an “inference-guiding role,” the gesture here can be

said to activate an “implicit pre-existing connection of a certain element (...) of experience to another” (Barcelona, 2009, p. 369). Gesturally triggered metonymic pathways of this nature may be seen as *natural inference schemata* (Panther and Thornburg, 2003, p. 8) or *vital relations* (Fauconnier and Turner, 2002: 93ff.): they intuitively draw on people’s embodied, situated ways of functioning not only in the physical world, but also in imaginary and/or abstract worlds (e.g., Sweetser, 2007, 2012).

For example, arriving at the contextualized meaning of the quick gestural indications in **Figure 3** relies on several cross-modal processes of pragmatic inferencing. Understanding these gestural portrayals as illustrating the WARM-WEATHER frame requires integrating the verbal utterance in (10) with information that is made visually salient. Apart from the iconic swimming gesture, the other frame elements mentioned in speech, such as the t-shirt and shorts, are evoked in a rather approximate way. In this multimodal portrayal, we can identify the following inferential pathways: two lead from the gesturing hands to the respective body parts and indicated items of clothing. Through the indexicality inherent to these gestures, what they allude to briefly constitutes a signifying, active zone (Langacker, 1987) that is profiled and thus perceptually foregrounded in this instance of multimodal meaning elaboration (see section “Experiential and Functional Domains”). Together these pathways lead more globally into the WARM-WEATHER frame, in the context of which these specific garments are commonly worn in combination. In these cases, but also more generally, the concurrent speech content is needed to disambiguate, via inferential processes, especially those gestures that only vaguely allude to something in the interlocutors’ environment or evolving discourse context.

So, although reference is one of metonymy’s chief functions, processes of pragmatic inferencing are often more crucially involved in assuring a gesture’s communicative function, at least from the perspective of the interpreter. Further gesture research is clearly needed to gain a fuller understanding of how speech and gesture interact in cross-modal processes of pragmatic inferencing including those that involve less accessible targets, for instance, through metaphoric construal (Mittelberg, 2006, 2017a).¹⁰ We will now look more closely at the junctures where such inferences tend to take place within visuo-kinetic signs.

Contiguity Relations Operationalized in Co-speech Gestures

From a semiotic perspective, similarity (iconicity), contiguity (indexicality), and conventionality (symbols, habits) constitute the three fundamental semiotic relations that may be established between a material sign carrier and what it signifies; in any given sign process, they typically mix to varying degrees (Peirce, 1960). The present proposal emphasizes that motivation in gesture relies on both similarity and contiguity and that both modes usually also interact with various pressures of conventionalization (e.g., Mittelberg, 2006, 2013, 2014). According to Peirce (1960), contiguity encompasses different kinds of factual connections,

¹⁰See also Calbris (2011: 78ff.) on *body-focused gestures* and Mittelberg and Waugh (2014) on *body part indices*.

notably physical impact, contact, and adjacency, as well as temporal and spatial proximity or distance. All of these may underpin indexical sign processes in which the material sign, for example, fingerprints left at a crime scene, points the interpreting mind toward the “object,” namely the person whose fingers caused traces of their impressions to adhere to surfaces through physical contact. Generally speaking, there are innumerable latent contiguity relations out there in the world, in our imagination, and in our embodied knowledge structures that may be operationalized when we are reasoning and communicating. This section will focus on contiguity relations that the speaker’s body forms with the physical or the imaginary world at her/his fingertips and that are intuitively drawn upon for multimodal meaning-making (cf. **Table 1** for an overview of the Peircean and Jakobsonian semiotic concepts discussed in this section).

Within cognitive linguistics, contiguity relations underpinning metonymic expressions are understood as either objectively given or cognitively construed (e.g., Panther and Radden, 1999; Dirven and Pörings, 2002). They are assumed to be contingent (Panther and Thornburg, 2003), that is to say, they may be canceled. These views are highly relevant to bodily semiotics and visuo-kinetic signs, for gesticulating hands typically do not manipulate real physical objects or surfaces; they only pretend to do so [as in simulating typing a paper, Example (8)]. In their prototype approach to conceptual contiguity and metonymy, Peirsman and Geeraerts (2006) posit the spatial and material domain as the prototypical core of contiguity. They present a continuum of strength of contact as the basis for spatial, temporal, as well as abstract domains (including events, actions, processes, and assemblies). For instance, in the

spatial/material domain, the continuum extends from spatial (i) part/whole (e.g., head/person) to less prototypical cases, such as (ii) containment/container (e.g., milk/glass), (iii) location/located (e.g., house/inhabitants), and (iv) entity/adjacent entity (e.g., person/clothing). The first is equivalent to a part-whole-frame metonymy. Reflecting diminishing degrees of strength of contact, the second captures the relationship between a bread basket and bread (as discussed in “Metonymy Underpins Schematic Gestural Patterns and Fully Codified Visuo-Kinetic Signs”), the third captures the relationship between a table and a client sitting at a table [as in Example (2)], and the last captures the connection between the speaker’s legs in Example (10) and the shorts she refers to verbally. We will now consider a view of contiguity that makes comparable distinctions, but places emphasis differently.

Jakobson’s (1956) account of contiguity relations has proven to be particularly suitable to describe the functions that metonymy may assume in experientially motivated gestural signs (Mittelberg, 2006, 2010, 2013; Mittelberg and Waugh, 2009, 2014). In his writings on aphasic disorders, Jakobson (1956) showed just how deeply rooted the distinction between similarity (iconicity/metaphor) and contiguity (indexicality/metonymy) is. Furthermore, he differentiated contiguity relations in the physical world, for example, between a knife and a fork, and those which combine items in a semiotic contexture, for example, linguistic units jointly forming a syntagm or a discourse (Waugh and Monville-Burston, 1990; Dirven and Pörings, 2002; Hopper and Traugott, 2003). Of particular relevance to understanding how metonymy is operationalized in gesture is Jakobson’s distinction between *inner contiguity* and *outer contiguity*. The following visual scene serves to illustrate these different operations, which will be applied to gesture below:

One must – and this is most important – delimit and carefully consider the essential difference between the two aspects of contiguity: the exterior aspect (metonymy proper), and the interior aspect (synecdoche, which is close to metonymy). To show the hands of a shepherd in poetry or the cinema is not the same as showing his hut or his herd [...]. The operation of synecdoche, with the part for the whole or the whole for the part, should be clearly distinguished from metonymic proximity. [...] the difference between inner and outer contiguity [...] marks the boundary between synecdoche and metonymy proper.
(Jakobson and Pomorska, 1983, p. 134).

TABLE 1 | Overview of semiotic foundations of metonymy in visuo-kinetic signs, drawing on Peircean semiotic theory and Jakobson’s view of metonymy as being derived from (outer and inner) contiguity and indexicality (as discussed in section “Contiguity Relations Operationalized in Co-speech Gestures”).

Semiotic relations (Peirce)	Semiotic modes (Peirce)	Metonymic principles grounded in contiguity and indexicality (Jakobson)
Similarity	Iconicity	Inner contiguity relations underpinning internal metonymy (e.g., partial, stylized iconic gesture standing for entire action involving more body parts and complex action routine)
Contiguity	Indexicality	Outer contiguity relations underpinning external metonymy (e.g., gesture for object involved in imitated action)
Conventionality habit	Symbolicity	Habitual/conventionalized metonymic operations (involving iconicity, indexicality, and symbolismity) <ul style="list-style-type: none">• Habitual actions and pragmatic inferences operationalizing inner/outer contiguity relations (e.g., engendering gestural patterns)• Coded metonymic operations drawing on inner/outer contiguity relations (e.g., underpinning sign language lexemes)

Inner Contiguity: Parts, Phases, Contours, and Essential Qualities
Inner contiguity underlies part-whole relationships, that is, between a part and another part, a part and the whole, or the whole and the part (Jakobson and Pomorska, 1983). *Internal metonymy* operationalizes these kinds of contiguity relations inherent to a given gestalt. For instance, in *everyone lives under one roof*, ‘roof’ evokes the entire house of which it constitutes a physical fragment. Hence, internal metonymy entails that the inner structure of a body, entity, or action is broken down into its component parts, phases, or any other characteristic, and that one of them is taken to imply a connected component or the entire gestalt structure.

In manual gestures and whole-body enactments, internal metonymy establishes a predominantly iconic ground for signification (Peirce, 1960; Sonesson, 2007; Mittelberg, 2013, 2014). That is, it relies upon a metonymic rendition of what it signifies based on a perceived or construed similarity. Internal metonymy may thus motivate processes of profiling and highlighting prototypical, or locally salient, aspects of a given, existent or imagined, experience or gestalt. For instance, a gesturally enacted onset, path, or manner of motion may evoke, in an abstracted and idealized manner, the corresponding, fully articulated physical action (e.g., the swimming gesture in **Figure 2**) or motion event (e.g., McNeill, 1992). It is via metonymy that iconic gestures may also give salience to contours, shapes, spatial dimensions, and other relevant qualities of objects, spaces, and other kinds of physical structures (Mittelberg and Waugh, 2014). In the study on transitive action gestures mentioned in Section “Experiential and Functional Domains” (Grandhi et al., 2011), an alternative way of enacting the apple-slicing scenario (**Figure 1**) was to use the hands as if they were the apple and the knife, respectively, rather than pretending to handle them. **Figure 4** shows two slightly different variants of this gestural technique, exemplifying the *representing* mode, according to Müller (1998, 2014).

- (11) *You need to slice the apple.*
 (12) *You need to cut the apple.*

In **Figure 4** [Examples (11) and (12)], both of the participants' hands exemplify the working of internal metonymy: The flat, vertically held hand looks and functions like the blade of a knife, that is, like the part of the kitchen tool would actually cut into an apple; the other, non-dominant hand forms a fist, thus resembling a round object, which, in this case, signifies an apple. Furthermore, the participant shown in **Figures 4B,C** opens up his hand, representing the apple, at the very moment when 'the knife' hits it, so that his fingers may be taken to iconically portray the apple slices resulting from the repeated cutting action in a schematic and partial fashion. In this visually effective instance of a gestural CAUSE-EFFECT metonymy, the semiotic affordances of the manual articulators are thus exploited to a great extent.

Outer Contiguity: In Touch With the Physical, Social, and Imagined World

Outer contiguity underlies metonymic expressions in which the profiled element is not part of, but externally contiguous and/or pragmatically related to the element that it enables an addressee to infer. *External metonymy* may draw on various kinds of outer contiguity relations and imply different degrees of metonymic proximity, such as contact, adjacency, impact, and cause/effect (Jakobson and Pomorska, 1983, p. 134). For instance, with respect to the metonymic source expressions “the ham sandwich” [Example (1)] and “Table 5” [Example (2)] referring to a restaurant client, the relevant contiguity relations hold between the client and the dish ordered earlier (temporal contiguity) and the table s/he is sitting at (spatial contiguity).

In gesture, contiguity holds between hands and the objects, tools, and surfaces with which speakers are (seemingly) in touch when communicating. Indexically anchoring the give and take of conversational exchanges in the actions of the human body, the material and social environment, or in imagined spaces (Sweetser, 2012; Mittelberg, 2017b), gestures readily (re-)establish and highlight such relations by instigating metonymic modes that operate at junctures of gesturing hands and contiguous persons or entities (as in the transitive action gesture in **Figure 1**).¹¹

In **Figure 5**, for instance, the purpose of the gestural enactment is not to iconically imitate someone who is holding something. While this is, via internal metonymy, the perceivable starting point of the enacted meaning construction, it is the imaginary entity, externally contiguous to the PUOH seemingly supporting it, that the speaker is verbally drawing attention to. In Example (13), the architecture student is describing an analogy between the architectural design process and a musical episode. Due to the basic physical action frame of holding a physical object, it is easy to infer a generic object. The latter here stands for an abstract concept, an analogy, via an embodied metonymic inference mechanism based on immediate metonymic proximity of the open palm and the imagined object. It is through action-based, metaphorical reification that the analogy becomes

¹¹See Mittelberg and Waugh (2014) for a typology of predominantly indexical or predominantly iconic gestures, metonymic chains triggered by gestures, and a continuum spanning varying degrees of metonymic proximity.



FIGURE 4 | Gestural examples of internal metonymy: the hands turn into an apple and a knife to portray how to slice an apple: the outspread fingers of the right hand in **(C)** also evoke the slices resulting from the cutting action. Video still **(A)** corresponds to Example (11); video stills **(B,C)** correspond to Example (12) (adapted from Grandhi et al., 2011). Written informed consent was obtained from the depicted individuals for the publication of this image.



FIGURE 5 | Gesture exemplifying an outer contiguity relationship between the left open flat hand and a virtual object seemingly placed on it. A muted index anchors a cross-modal process of pragmatic inferencing: the abstract discourse content, an analogy metaphorically construed as a tangible entity, can be inferred from the surface of the open hand via external metonymy (Example 13). Written informed consent was obtained from the depicted individuals for the publication of this image.

a tangible and thus intersubjectively sharable element in the discourse context (Mittelberg, 2008; Mittelberg and Joue, 2017).

- (13) *Es gab ja die Analogie zur... zur Musik, also... oder... oder auch zu 'ner Interpretation*
(‘There was the analogy to... to music, so... or... or even to an interpretation’).

Under the present view, it is through a cross-modal process of pragmatic inferencing and a low degree of indexicality (Mittelberg, 2017a,c) that the PUOH is pointing to the existence and relevance of the verbally referenced analogy. Concurrently, the speaker’s right hand with the palm turned downward is oriented toward his left hand and thus creates an additional index leading to the analogy. This gesture thus heightens the relevance of an idea that seems to be physically graspable through the ongoing multimodal description.

These observations further support the idea that the source meaning, embodied in the form of a hand configuration and/or movement, remains present and perceptually salient in metonymic mappings, while the target meaning, that is, the discourse contents (such as the analogy in **Figure 5**) is cognitively prominent in the ongoing exposition (e.g., Panther and Thornburg, 2004, p. 95, 105; Mittelberg, 2006). We can also say that such object-oriented action gestures may trigger a frame-internal metonymic shift at outer contiguity junctions constituted by the hands and the virtual objects and tools they seem to be holding or manipulating (Mittelberg and Waugh, 2014). For example, the apple-slicing scenario in **Figure 1** exemplifies the underlying metonymic mapping ACTION-FOR-OBJECT INVOLVED IN ACTION; in addition, the gesture in **Figure 5** simultaneously relies on the relation PRESENTATION-FOR-PRESENTED (e.g., Panther and Radden, 1999; Panther and Thornburg, 2003; Barcelona, 2009; see also Mittelberg and Joue,

2017 on *gestural framing actions*). In both cases, what we actually see are the physical actions, but through following the linguistic cues in the unfolding discourse our attention shifts to the implied items and ideas.

A recent study combining behavioral and brain-imaging experiments (Joue et al., 2018) provides some initial neuroscientific evidence for processing differences that seem to broadly reflect the metonymic principles distinguished by Jakobson and Pomorska (1983) and discussed in this section. The study participants were shown video recordings of persons verbally describing and gesturally performing actions in which an object/tool in question was either represented by a finger/hand (internal metonymy or *body-part-as-object*) or the person was pretending to be holding or otherwise manipulating an object or tool (external metonymy or *pantomime*; see, e.g., Lausberg et al., 2003). Results suggest that metonymy may guide an interpreting mind to focus primarily on either locally relevant features (part-for-whole metonymy) or more globally relevant aspects (frame metonymy) of what is being communicated (Joue et al., 2018; see also Grandhi et al., 2011 on a user study showing clear preferences for external metonymy).

Experiential, Metonymic Bases for Metaphoricity in Gesture

Metonymy and metaphor have been found to interact to varying degrees in language and other multimodal forms of communication (e.g., Jakobson, 1956; Goossens, 1990; Barcelona, 2000a; Radden, 2000; Mittelberg, 2002, 2008; Panther et al., 2009; Benczes et al., 2011; Kövecses, 2013; Littlemore, 2015; Hampe, 2017; Ruiz de Mendoza Ibanez, 2017). Investigating how indexical and iconic principles jointly guide the interpretation of predominantly metaphoric gestures, Mittelberg and Waugh (2009) suggest two distinct but intertwined semiotic processes in which metonymy leads into metaphor. For example, to reconstruct the meaning of the gesture evoking an analogy in **Figure 5**, we can first assume a process of metonymic inferencing as described in Sections “Reference and Pragmatic Inferencing in Gesture” and “Contiguity Relations Operationalized in Co-speech Gestures.” The metonymic source, namely the flat open hand involved in the source action (Mittelberg and Joue, 2017) of holding something, points to the adjacent metonymic target: the virtual object involved in the action. Second, the same imaginary object serves as the source of the metaphoric mapping whose target is the abstract notion of analogy referred to verbally (see also Taub, 2001 and Meir, 2010 on *double mappings* in sign language). Note that in this example of a gesturally enacted metaphor, the concurrent speech is non-figurative (see also Cienki and Müller, 2008; Mittelberg, 2008, 2014).

Furthermore, the gesture in **Figure 5** is an instance of a gesturally expressed *primary metaphor* (Grady, 1997; Hampe, 2017), namely IDEAS ARE OBJECTS. It thereby evokes the basic physical action and object frame of handling objects (Mittelberg, 2017a), which involves a primary scene (Grady, 1997), a prototypical event (Slobin, 1985), and a scene basic to human experience (Goldberg, 1995; as discussed in Section “Gestural Frame Evocation at Varying Levels of Groundedness and Complexity”). It is hence central to the present perspective

on multimodal metonymy that “(f)rame metonymy is closely tied to the kind of correlations which are involved in experientially based metaphors, in particular Primary Metaphors (...). It is precisely the development of a complex frame out of a correlated simpler frame which makes a primary scene so powerful” (Dancygier and Sweetser, 2014, p. 137). We can draw from these insights that metaphoricity in gesture needs to be analyzed in view of its experientially grounded, metonymic bases, which may be predominantly iconic or predominantly indexical (e.g., Mittelberg and Waugh, 2014). This also further supports the idea that metonymy is experientially more basic than metaphor. (Cf. **Table 2** for an overview of the different approaches to metonymy discussed in this section).

METONYMY UNDERPINS SCHEMATIC GESTURAL PATTERNS AND FULLY CODED VISUO-KINETIC SIGNS

Throughout the foregoing discussion, we have seen how metonymic modes may motivate various processes of experientially grounded abstraction and schematization

with respect to particular gestures. We will now consider how metonymy may be said to also underpin the emergence of gestural patterns (in section “Enacted Schematicity: Pragmatically Driven Patterns in Co-speech Gesture”) as well as fully coded visuo-kinetic signs (in section “Metonymic Principles Operating in Signed Languages”).

Although gestures and signed languages largely share the same articulators and space as a medium of articulation, they also differ in the ways in which they are ‘visual’ and act as signs (e.g., Liddell, 2003; Wilcox, 2004c; Sweetser, 2009; Perniss et al., 2010; Kendon, 2014; Müller, 2017). In many discourse contexts, spontaneous gestures can afford to be quite allusive, idiosyncratically reduced semiotic gestalts, for they do not need to fulfill well-formedness conditions in the way that emblems and linguistic symbols in signed languages do. Gestures may in fact push metonymic form reduction and schematization to quite extreme degrees. This is partly because, most of the time, gestures do not carry the full load of meaning-making: The concurrent spoken utterance gives them a hand, so to speak, thus disambiguating potentially polysemous hand shapes and movements (e.g., Müller, 1998; Calbris, 2011).

A central point that this paper wishes to make is that using the umbrella term *visuo-kinetic signs* to encompass

TABLE 2 | Overview of experiential and cognitive bases and the corresponding metonymic operations in accordance with cognitive linguistic approaches to metonymy.

Experiential and conceptual bases	METONYMY	Metaphor
Domains (Section “Experiential and Functional Domains”)	Metonymic operations profiling certain elements/qualities within the same <ul style="list-style-type: none"> • <i>Experiential/functional domain or domain matrix</i> • <i>Idealized Cognitive Model (ICM) or cultural model</i> 	Crossmodal metonymy-metaphor interaction:
Frames and scenes (Section “Semantic Frames and Familiar Scenes of Experience”)	<ul style="list-style-type: none"> • <i>Frame metonymy and part-for-whole frame metonymy</i> • Metonymic organization of semantic/syntactic frames 	Correlations between profiled parts across experiential domains/frames, e.g.,:
Gestural frame evocation (Section “Gestural Frame Evocation at Varying Levels of Groundedness and Complexity”)	Gestures metonymically profile salient elements of scenes and frame structures at varying levels of groundedness, schematicity, and complexity: <ul style="list-style-type: none"> • <i>Basic object/action frames</i> • <i>intermediately complex frames</i> • <i>highly complex and abstract frame structures</i> 	Embodied metonymic bases of metaphoricity: <ol style="list-style-type: none"> Gestural action evokes the object involved in the action via metonymy; The same imagined tangible object signifies an abstract idea via metaphor.
Reference/inference (Section “Reference and Pragmatic Inferencing in Gesture”)	<ul style="list-style-type: none"> • Referential metonymy and indirect reference • Pragmatic inferencing guided by embodied habits of action and interpretation and/or linguistic cues 	(see section “Experiential, Metonymic Bases for Metaphoricity in Gesture”)
Contiguity (Section “Contiguity Relations Operationalized in Co-speech Gestures”) (Table 1)	Metonymic correlations based on physical/conceptual contiguity. Contiguity relations get operationalized via predominantly iconic or indexical modes in gesture and/or linguistic cues: <ul style="list-style-type: none"> • <i>Inner contiguity</i> (internal metonymy; synecdoche), e.g., profiling quality or phase of an action • <i>Outer contiguity</i> (external metonymy), e.g., between hand and object involved in action 	Interacting metonymic and metaphoric processes may propel <ul style="list-style-type: none"> • Pragmatic functions • Schematization • Grammaticalization (see section “Enacted Schematicity: Pragmatically Driven Patterns in Co-speech Gesture”)

In addition, the table includes an application of semantic frames to gestures, devising varying levels of groundedness, schematicity, and complexity in gestural frame evocation, as well as aspects of metonymy-metaphor interaction in gesture.

co-speech gestures and signed languages allows us to elucidate some commonalities regarding certain core principles of metonymically driven sign constitution and interpretation.

Enacted Schematicity: Pragmatically Driven Patterns in Co-speech Gesture

A central goal in gesture research has been to identify patterns in gestural practices within and across individual speakers, languages, discourses, contexts, communities, and cultures (e.g., Kendon, 2004; McNeill, 2005; Streeck et al., 2011). Certain co-speech gestures have indeed been found to exhibit relatively high degrees of patterning and conventionality. Under the present view on regularities in gesture, *conventionality* strongly pertains to the Peirce (1960) notion of *habit*, rather than to imposed, *symbolic* codes in the narrow sense of the term (Mittelberg, 2006). Highly frequent and routinized gestures, particularly those that (also) fulfill pragmatic functions, indeed show an increased ‘visibility’ in multimodal interaction: for example, *gesture families* (e.g., Kendon, 2004); *recurrent gestures* such as the PUOH (Müller, 2004, 2017) or the *cyclic gesture* (Ladewig, 2011, 2014; see also Bressemer, 2014; Bressemer and Müller, 2014); and/or gestures enacting embodied image and force schemas (e.g., Mittelberg, 2008, 2018; Cienki, 2013; Wehling, 2017).¹²

Suggesting that scenes basic to human experience (Goldberg, 1995, p. 5) may underpin entrenched patterns in both language and gesture, it was argued in Section “Toward a Frame-Based Account of Embodied Metonymy in Gesture” (drawing on Mittelberg, 2017a and Mittelberg and Joue, 2017) that certain gestures tend to metonymically profile salient aspects of deeply embodied, routinized aspects of scenes, that is, the motivating context of semantic frames (Fillmore, 1977, 1982). The next logical steps of this rationale involve examining how metonymy conditions gradual, pragmatically motivated processes of grammaticalization (Hopper, 1998; Hopper and Traugott, 2003; Bybee, 2010) in gesture, how the resulting schematic gestures evoke correlated syntactic frames (e.g., Goldberg, 1995, 1998), and how they may partake in multimodally instantiated constructions (e.g., Mittelberg, 2017c; see also contributions in Zima and Bergs, 2017). A full account of these complex phenomena cannot be provided here, but we will see an example of how metonymy factors into gestural schematicity below.

Regarding the meaning of constructions in language, Barcelona (2009) ascribes a fundamental role to metonymy and pragmatic inferences (see also Panther et al., 2009). According to the present, admittedly preliminary, consideration of comparable processes in gesture, habituated physical actions and repeated similar acts of gesturing involve metonymy through propelling the establishment of not only individual, metonymically reduced gestures, but also more schematic gestural patterns, notably via discourse-driven routinization (e.g., Haiman, 1994; Hopper and Traugott, 2003) of certain physical actions. Such commonly used, more strongly conventionalized, visuo-kinetic signs

should evidence the metonymic processes discussed in this paper to high degrees. Gestures displaying this increased level of *embodied schematicity* are likely to combine referential (including metaphoric) and pragmatic functions, and their interpretation can be expected to rely on entrenched processes of pragmatic inferencing (such as the ones described in section “Reference and Pragmatic Inferencing in Gesture”).¹³

For instance, basic manual actions, such as holding or giving something to someone, have been shown to entail schematic scenes that underpin prototypical cases of transitive or ditransitive argument structure in language (Goldberg, 1995). In German, the full verb *geben* (give), a three-place predicate, underwent a process of grammaticalization engendering the existential construction *es gibt* ‘it gives’ (there is/are; Newman, 1998). In a recent study on multimodal instantiations of this impersonal construction (Mittelberg, 2017c), it has been argued that the manual action of giving also serves as *experiential substrate* in processes of embodied grammaticalization that result in gestural existential markers observed to co-occur with *es gibt*. These gestural markers tend to enact reduced and more schematic variants of the full action of giving. To illustrate this point, let us revisit the PUOH gesture in **Figure 5** (discussed in section “Outer Contiguity: In Touch With the Physical, Social, and Imagined World”). In his left hand, the participant is seemingly holding the analogy he is talking about while using an *es gibt* construction to refer to it verbally (Example 13). The basic sense of the full verb *geben* (give) as well as the basic scene it evokes still resonate not only in this intransitive linguistic construction, but also in the gestural enactment that co-occurred with it. In essence, metonymic reduction can be said to motivate such frequently occurring, schematic communicative gestures out of fully fledged, object-oriented physical actions that originally involve object transfer. The act of giving is reduced to an act of unimanual holding that exhibits a decreased degree of transitivity and iconicity, thus evoking, for instance, a scene of existence, or presence, rather than a scene of object transfer.

These grammaticalized gestural markers of existence tend to stay rather close to the speaker’s body instead of reaching toward an (imagined) receiver. The hands are also more relaxed and reveal less effort than would be necessary to actually hold something. In addition, these visuo-kinetic existential markers tend to combine referential dimensions, afforded through metonymy, with modal or epistemic, that is, pragmatic functions (e.g., Sweetser, 1990). They also tend to express subjective and interactive dimensions of meaning (e.g., Hopper and Traugott, 2003), such as in Example 13, where the speaker points out something that seems obvious to him (for further details and examples see Mittelberg, 2017c; see also Bressemer and Müller, 2014; Müller, 2017). Reduced degrees of iconicity and indexicality seem to push such commonly used gestures closer toward the juncture of *habit-driven, embodied grammaticalization* and gesture pragmatics. Further research is clearly needed to establish how these initial insights play out across speakers, languages, and discourse contexts.

¹²Pointing gestures (e.g., McNeill, 1992; Haviland, 2000; Kendon, 2004; Fricke, 2007) are also commonly observed gestural practices, but limits of space do not allow them to be included in the present discussion.

¹³For a brain-imaging study on perceived conventionality in co-speech gestures see Wolf et al. (2017).

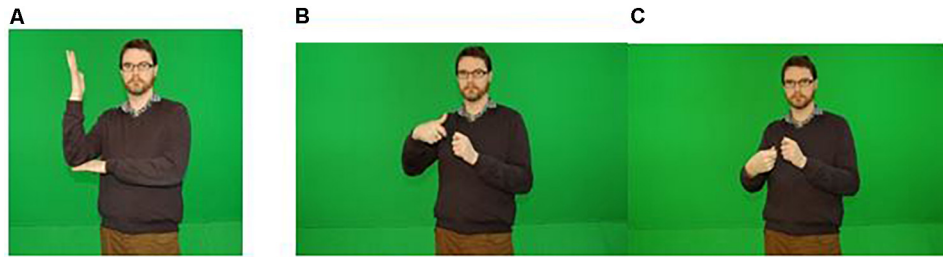


FIGURE 6 | Photo (A) shows the DGS sign for 'Baurrr' (tree) exemplifying *internal metonymy*; photos (B,C) show the DGS sign for 'Banana,' exemplifying *external metonymy*. Written informed consent was obtained from the depicted individuals for the publication of this image.

The observations discussed here are akin to work on grammaticalization in signed languages, in the context of which gestures have been shown to serve as the *substrate* of certain lexical and/or grammatical signs (Janzen and Shaffer, 2002). Bearing this in mind, we will now turn to how metonymy operates in signed language.

Metonymic Principles Operating in Signed Languages

Metonymy has been ascribed an important role in the construction of form and meaning in signed languages, for example, in ASL (e.g., Mandel, 1977; Taub, 2001; Liddell, 2003; Wilcox et al., 2003; Wilcox, 2004a,b), French Sign Language (LSF; e.g., Bouvet, 1997), German Sign Language (DGS; e.g., Kutscher and Lincke, 2012); and Israeli Sign Language (ISL; e.g., Meir, 2010; Meir and Cohen, 2018 *fc.*). For instance, investigating how iconicity and metaphor interact in ASL, Taub (2001) suggests a set of principles of sign constitution including *image selection*, *schematization*, and *encoding*. Metonymy particularly comes into play at the image selection stage: The ASL sign for 'academic degree', for instance, consists in showing the gestalt and length of a rolled-up diploma shaped like a cylinder. The sign portrays a tangible element that is pragmatically correlated with the target meaning within the same frame: "The degree itself is a non-physical title, rather than a physical object, and so a salient object is chosen for the purposes of creating an iconic sign" (Taub, 2001, p. 46).

Let us now see how internal and external metonymy (as introduced in section "Contiguity Relations Operationalized in Co-speech Gestures") are manifested in DGS. In **Figure 6A**, the lexical sign for 'Baum' (tree) exemplifies the workings of internal metonymy in the form of a bimanually achieved schematic icon that profiles the salient, structural parts of a tree, namely its trunk and branches, as well as the ground in which it is rooted. By contrast, the DGS sign for 'Banane' (banana) is a good example of how outer contiguity relations between the hands and a manipulated object are drawn upon to pragmatically infer what is signified (**Figures 6B,C**). While internal metonymy underpins the iconic hand shapes and movements as such, it is via external (or frame) metonymy that the hands' shapes and actions evoke the implied (invisible) fruit. Although the peeling action is physically made salient, it is the object undergoing the action that is being referred to via this iconic visuo-kinetic lexeme. The

latter may be said to evoke a basic action and object frame as discussed in section "Reference and Pragmatic Inferencing in Gesture" (Mittelberg, 2017a), involving the metonymic mapping ACTION-FOR-OBJECT INVOLVED IN ACTION (see section "Toward a Frame-Based Account of Embodied Metonymy in Gesture"; see also Wilcox et al., 2003).

CONCLUDING REMARKS

The insights offered in the foregoing discussion provide further support for the idea that metonymy is a fundamental principle that operates across different modalities of experience, thought, and expression. The chief goal of this paper was to characterize and evidence the inherently metonymic nature of co-speech gestures. Combining cognitive linguistic and semiotic perspectives on how embodied metonymic principles may underpin the formation and interpretation of gestures, the discussion has shown how a frame-based account may integrate related concepts such as scenes, experiential domains, contiguity (indexicality), similarity (iconicity), and habit/conventionality (symbolicity). Under this unified view, these different concepts provide an insightful lens onto various experientially grounded processes of metonymic motivation that tend to pragmatically induce, in one way or another, not only the forms and functions, but also the habit-driven processes of patterning and schematization that are discernable in gesture. We also saw gestural evidence for the claim that metonymy is experientially more basic than metaphor and hence often feeds into correlated metaphoric processes.

How metonymy plays out in signed language could only be briefly touched upon here in comparison to its role in co-speech gesture. We can preliminarily conclude that metonymic processes typically apply on-the-fly and from scratch in gestures, whose forms and potential meanings are highly context-dependent and not as strongly stabilized as they are in signed languages. However, what I am suggesting here is that the set of metonymic principles discussed in this paper seem to generally operate in visuo-kinetic signs, thus engendering similarly principled ways of forming embodied signs, as well as guiding inferential processes that are implied in their interpretation. Exactly how these metonymic mechanisms systematically compare and differ in gesture and signed language, including gestures occurring within

signed discourse, needs to be established through future research across languages, modalities, and discourse genres. Empirical investigation into how metonymic processes are conditioned by interacting experiential, physical, cognitive, cross-modal, modality-specific, discourse, interactional, and cultural forces will no doubt further our understanding of the complex dynamics of multimodal face-to-face interaction.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

- Arnheim, R. (1969). *Visual Thinking*. Oakland, CA: University of California Press.
- Barcelona, A. (ed.). (2000a). *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*. Berlin: Mouton de Gruyter.
- Barcelona, A. (2000b). "On the plausibility of claiming a metonymic motivation of conceptual metaphor," in *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*, ed. A. Barcelona (Berlin: Mouton de Gruyter), 31–58.
- Barcelona, A. (2003). "The case of a metonymic basis of pragmatic inferencing: evidence from jokes and funny anecdotes," in *Metonymy and Pragmatic Inferencing*, eds K.-U. Panther and L. L. Thornburg (Amsterdam: John Benjamins), 81–102. doi: 10.1075/pbns.113.07bar
- Barcelona, A. (2009). "Motivation of construction meaning and form: the roles of metonymy and inference," in *Metonymy and Metaphor in Grammar*, eds K.-U. Panther, L. L. Thornburg, and A. Barcelona (Amsterdam: John Benjamins), 363–401.
- Bavelas, J. B., Chovil, N., Coates, J. L., and Roe, L. (1995). Gestures specialized for dialogue. *Pers. Soc. Psychol. Bull.* 21, 394–405. doi: 10.1177/0146167295214010
- Benczes, R., Barcelona, A., and Ruiz de Mendoza Ibáñez, F. J. (eds). (2011). *Defining Metonymy in Cognitive Linguistics: Towards a Consensus View*. Amsterdam: John Benjamins. doi: 10.1075/hcp.28
- Bouvet, D. (1997). *Le Corps et la Métaphore dans les Langues Gestuelles: A la Recherche des Modes de Production des Signes*. Paris: L'Harmattan.
- Bouvet, D. (2001). *La Dimension Corporelle de la Parole: Les Marques Posturo-Mimo-Gestuelles de la Parole, Leurs Aspects Métonymiques et Métaphoriques, et Leur Rôle au Cours d'un Récit*. Paris: Peeters.
- Bressem, J. (2014). "Repetitions in gesture," in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1037–1958.
- Bressem, J., and Müller, C. (2014). "A repertoire of German recurrent gestures with pragmatic functions," in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Berlin: De Gruyter Mouton), 1575–1591.
- Burke, K. (1941). Four master tropes. *Kenyon Rev.* 3, 421–438. doi: 10.1080/10410236.2013.835782
- Bybee, J. L. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511750526
- Calbris, G. (1990). *The Semiotics of French Gestures*. Bloomington, IN: The University of Indiana Press.
- Calbris, G. (2011). *Elements of Meaning in Gesture*. Amsterdam: John Benjamins. doi: 10.1075/gs.5
- Cienki, A. (1998). "Metaphoric gestures and some of their relations to verbal metaphoric expressions," in *Discourse and Cognition: Bridging the Gap*, ed. J.-P. Koenig (Stanford, CA: CSLI Publications), 189–204.
- Cienki, A. (2013). Mimetic schemas and image schemas in cognitive linguistics and gesture studies. *Rev. Cogn. Linguist.* 11, 417–441. doi: 10.1075/rcl.11.2.13cie
- Cienki, A., and Müller, C. (eds). (2008). *Metaphor and Gesture*. Amsterdam: John Benjamins. doi: 10.1075/gs.3

ACKNOWLEDGMENTS

The author wishes to thank the editors of this research topic on "visual language", Wendy Sandler and Marianne Gullberg, as well as three reviewers for their helpful comments and suggestions. Much appreciation is extended to Linda Waugh for co-developing ideas on metonymy in gesture in earlier stages of this line of research, as well as to Eve Sweetser for continuously providing insightful feedback on the ideas brought together in this paper. Special thanks also go to Mary M. Copple for helpful feedback on the manuscript. The research reported on in this article was supported by the Excellence Initiative of the German Federal and State Government.

- Coulson, S. (2001). *Semantic Leaps. Frame-Shifting and Conceptual Blending in Meaning Construction*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9780511551352
- Croft, W. (1993). The role of domains in the interpretation of metaphor and metonymy. *Cogn. Linguist.* 4, 335–370. doi: 10.1515/cogl.1993.4.4.335
- Dancygier, B., and Sweetser, E. E. (2014). *Figurative Language*. Cambridge: Cambridge University Press.
- Dirven, R., and Pörrings, R. (eds). (2002). *Metaphor and Metonymy in Comparison and Contrast*. Berlin: De Gruyter Mouton.
- Dudis, P. (2004). Body partitioning and real-space blends. *Cogn. Linguist.* 15, 223–238. doi: 10.1515/cogl.2004.009
- Enfield, N. (2009). *The Anatomy of Meaning: Speech, Gestures, and Composite Utterances*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511576737
- Fauconnier, G., and Turner, M. (2002). *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York, NY: Basic Books.
- Fillmore, C. J. (1975). "An alternative to checklist theories of meaning," in *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society* Cathy Cogen, eds H. Thompson, G. Thurgood, K. Whistler, and J. Wright (Berkeley, CA: Berkeley Linguistics Society), 123–131.
- Fillmore, C. J. (1977). "Scenes-and-frames semantics," in *Linguistic Structure Processing*, Vol. 4, ed. A. Zampolli (Amsterdam: North Holland), 55–81.
- Fillmore, C. J. (1982). "Frame semantics," in *Linguistics in the Morning Calm*, ed. Linguistic Society of Korea (Seoul: Hanshin), 111–137.
- Forceville, C. J. (2007). "Metonymy in visual and audiovisual discourse," in *The World Told and the World Shown: Multisemiotic Issues*, eds E. Ventola and A. J. Moya Guijarro (London: Palgrave Macmillan), 56–74.
- Forceville, C. J., and Urios-Aparisi, E. (eds). (2009). *Multimodal Metaphor*. Berlin: De Gruyter Mouton. doi: 10.1515/9783110215366
- Fricke, E. (2007). *Origo, Geste und Raum – Lokaldeixis im Deutschen*. Berlin: Walter de Gruyter.
- Fricke, E. (2012). *Grammatik Multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin: De Gruyter Mouton. doi: 10.1515/9783110218893
- Gibbs, R. W. Jr. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge: Cambridge University Press.
- Gibbs, R. W. Jr. (1999). "Speaking and thinking with metonymy," in *Metonymy in Language and Thought*, eds G. Radden and Z. Kövecses (Amsterdam: John Benjamins), 61–76. doi: 10.1075/hcp.4.04gib
- Gibbs, R. W. Jr. (2006). *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.
- Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: The University of Chicago Press.
- Goldberg, A. E. (1998). "Patterns of experience in patterns of language," in *The New Psychology of Language*, ed. M. Tomasello (London: Lawrence Erlbaum Assoc.), 203–219.
- Goodwin, C. (2011). "Contextures of action," in *Embodied Interaction: Language and Body in the Material World*, eds J. Streeck, C. Goodwin, and C. LeBaron (Cambridge: Cambridge University Press), 182–193.
- Goossens, L. (1990). Metaphtonymy: the interaction of metaphor and metonymy in expressions for linguistic action. *Cogn. Linguist.* 1, 323–340. doi: 10.1515/cogl.1990.1.3.323

- Grady, J. (1997). *Foundations of Meaning: Primary Metaphors and Primary Scenes*. Ph.D. dissertation, University of California, Berkeley, CA.
- Grandhi, S. A., Joue, G., and Mittelberg, I. (2011). "Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces," in *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, Atlanta, GA, 821–824. doi: 10.1145/1978942.1979061
- Haiman, J. (1994). "Ritualization and the development of language," in *Perspectives on Grammaticalization*, ed. W. Pagliuca (Amsterdam: John Benjamins), 3–29. doi: 10.1075/cilt.109.07hai
- Hampe, B. (ed.). (2017). *Metaphor: Embodied Cognition and Discourse*. Cambridge: Cambridge University Press. doi: 10.1017/9781108182324
- Hassemer, J. (2015). *Towards a Theory of Gesture Form Analysis: Principles of Gesture Conceptualization, with Empirical Support from Motion-Capture Data*. Ph.D. dissertation, RWTH Aachen University, Aachen.
- Haviland, J. (2000). "Pointing, gesture spaces, and mental maps," in *Language and Gesture*, ed D. McNeill (Cambridge: Cambridge University Press), 13–46.
- Hopper, P. (1998). "Emergent grammar," in *The New Psychology of Language*, ed. M. Tomasello (London: Lawrence Erlbaum Assoc), 155–175.
- Hopper, P., and Traugott, E. C. (2003). *Grammaticalization*, 2nd Edn. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139165525
- Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychon. Bull. Rev.* 15, 495–514. doi: 10.3758/PBR.15.3.495
- Hutchins, E. (2005). Material anchors for conceptual blends. *J. Pragmat.* 37, 1555–1577. doi: 10.1016/j.pragma.2004.06.008
- Jakobson, R. (1956). "Two aspects of language and two types of aphasic disturbances," in *Roman Jakobson – On Language*, eds L. R. Waugh and M. Monville-Burstion (Cambridge, MA: Belknap of Harvard University Press), 115–133.
- Jakobson, R. (1972). Motor signs for 'yes' and 'no'. *Lang. Soc.* 1, 91–96. doi: 10.1017/S0047404500006564
- Jakobson, R., and Pomorska, K. (1983). *Dialogues*. Cambridge: Harvard University Press.
- Janzen, T., and Shaffer, B. (2002). "Gesture as substrate in the process of ASL grammaticalization," in *Modality and Structure in Signed and Spoken Language*, eds R. P. Meier, K. Cormier, and D. Quintos-Pozos (Cambridge: Cambridge University Press), 199–223.
- Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226177847.001.0001
- Joue, G., Boven, L., Willmes, K., Evola, V., Demenescuz, L., Hassemer, J., et al. (2018). Handling or being the concept: an fMRI-study on metonymy representations in verbal gestures. *Neuropsychologia* 109, 232–244. doi: 10.1016/j.neuropsychologia.2017.12.033
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572
- Kendon, A. (2014). Semiotic diversity in utterance production and the concept of 'language'. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130293. doi: 10.1098/rstb.2013.0293
- Kövecses, Z. (2013). The metaphor–metonymy relationship: correlation metaphors are based on metonymy. *Metaphor Symb.* 28, 75–88. doi: 10.1080/10926488.2013.768498
- Kutscher, S., and Lincke, E.-S. (2012). "Motivated sign formation in hieroglyphic Egyptian and German sign language (DGS)," in *Lexical Semantics in Ancient Egyptian*, eds E. Grossman, S. Polis, and J. Winand (Göttingen: Seminar für Ägyptologie und Koptologie), 113–140.
- Ladewig, S. H. (2011). *Putting a Recurrent Gesture on its Cognitive Basis*. *CogniTextes* 6. Available at: <http://cognitextes.revues.org/406> [accessed 28 June 2013].
- Ladewig, S. H. (2014). "Recurrent gestures," in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressen (Berlin: De Gruyter Mouton), 1558–1574.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago, IL: The University of Chicago Press. doi: 10.7208/chicago/9780226471013.001.0001
- Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: Chicago University Press.
- Langacker, R. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
- Langacker, R. (2009). "Metonymic grammar," in *Metonymy and Metaphor in Grammar*, eds K.-U. Panther, L. L. Thornburg, and A. Barcelona (Amsterdam: John Benjamins), 5–71.
- Lausberg, H., Cruz, R. F., Kita, S., Zaidel, E., and Pfito, A. (2003). Pantomime to visual presentation of objects: left hand dyspraxia in patients with complete callosotomy. *Brain* 126, 343–360. doi: 10.1093/brain/awg042
- Liddell, S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511615054
- Littlemore, J. (2015). *Metonymy: Hidden Shortcuts in Language, Thought and Communication*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107338814
- Mandel, M. (1977). "Iconic devices in American sign language," in *On the Other Hand: New Perspectives on American Sign Language*, ed. L. Friedman (New York, NY: Academic Press), 57–107.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: Chicago University Press.
- McNeill, D. (2005). *Gesture and Thought*. Chicago, IL: Chicago University Press. doi: 10.7208/chicago/9780226514642.001.0001
- Meir, I. (2010). Iconicity and metaphor: constraints on metaphorical extension of iconic forms. *Language* 86, 865–896. doi: 10.1353/lan.2010.0044
- Meir, I., and Cohen, A. (2018). Metaphor in sign language. *Front. Psychol.* 9:1025. doi: 10.3389/fpsyg.2018.01025
- Merleau-Ponty, M. (1962). *Phenomenology of Perception*. New York: Humanities Press.
- Mittelberg, I. (2002). The visual memory of grammar: iconographical and metaphorical insights. *Metaphorik.de* 69–89.
- Mittelberg, I. (2006). *Metaphor and Metonymy in Language and Gesture: Discourse Evidence for Multimodal Models of Grammar*. Ph.D. Dissertation, Cornell University, Ann Arbor, MI.
- Mittelberg, I. (2008). "Peircean semiotics meets conceptual metaphor: iconic modes in gestural representations of grammar," in *Metaphor and Gesture*, eds A. Cienki and C. Müller (Amsterdam: John Benjamins), 115–154.
- Mittelberg, I. (2010). Interne und externe Metonymie: Jakobsonsche Kontiguitätsbeziehungen in redebegleitenden Gesten. *Spr. Lit.* 41, 112–143. doi: 10.1163/25890859-041-01-90000007
- Mittelberg, I. (2013). "The embodied mind: cognitive-semiotic principles as motivating forces in gesture," in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressen (Berlin: De Gruyter Mouton), 755–785.
- Mittelberg, I. (2014). "Gestures and iconicity," in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressen (Berlin: De Gruyter Mouton), 1712–1732.
- Mittelberg, I. (2017a). Embodied frames and scenes: body-based metonymy and pragmatic inferencing in gesture. *Gesture* 16, 204–233. doi: 10.1075/gest.16.2.03mit
- Mittelberg, I. (2017b). Experiencing and construing spatial artifacts from within: simulated artifact immersion as a multimodal viewpoint strategy. *Cogn. Linguist.* 28, 381–415. doi: 10.1515/cog-2016-0124
- Mittelberg, I. (2017c). Multimodal existential constructions in German: manual actions of giving as experiential substrate of grammatical and gestural patterns. *Linguist. Vanguard* 3:s1. doi: 10.1515/lingvan-2016-0047
- Mittelberg, I. (2018). Gestures as image schemas and force gestalten: a dynamic systems approach augmented with motion-capture data analyses. *Cogn. Semiotics* 11:20180002. doi: 10.1515/cogsem-2018-0002
- Mittelberg, I., and Joue, G. (2017). "Source actions ground metaphor via metonymy: towards a frame-based account of gestural action in multimodal discourse," in *Metaphor: Embodied Cognition and Discourse*, ed. B. Hampe (Cambridge: Cambridge University Press), 119–137.
- Mittelberg, I., and Waugh, L. R. (2009). "Metonymy first, metaphor second: a cognitive-semiotic approach to multimodal figures of thought in co-speech

- gesture,” in *Multimodal Metaphor*, eds C. Forceville and E. Urios-Aparisi (Berlin: De Gruyter Mouton), 329–356.
- Mittelberg, I., and Waugh, L. R. (2014). “Gestures and metonymy,” in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 1747–1766.
- Müller, C. (1998). *Redebegleitende Gesten: Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Spitz Verlag.
- Müller, C. (2004). “Forms and uses of the palm up open hand: a case of a gesture family?,” in *The Semantics and Pragmatics of Everyday Gesture: Proceedings of the Berlin Conference April 1998*, eds C. Müller and R. Posner (Berlin: Weidler Verlag), 233–256.
- Müller, C. (2014). “Gestural modes of representation as techniques of depiction,” in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction* 38.2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: De Gruyter Mouton), 1687–1702.
- Müller, C. (2017). How recurrent gestures mean: conventionalized contexts-of-use and embodied motivation. *Gesture* 16, 277–304. doi: 10.1075/gest.16.2.05mul
- Nerlich, B., and Clarke, D. D. (2001). Ambiguities we live by: towards a pragmatics of polysemy. *J. Pragmat.* 33, 1–20. doi: 10.1016/S0378-2166(99)00132-0
- Newman, J. (1998). “The origin of the German es gibt construction,” in *The Linguistics of Giving*, ed. J. Newman (Amsterdam: John Benjamins), 307–325.
- Panther, K. U., and Radden, G. (eds). (1999). *Metonymy in Language and Thought*. Amsterdam: John Benjamins. doi: 10.1075/hcp.4
- Panther, K. U., and Thornburg, L. L. (eds). (2003). *Metonymy and Pragmatic Inferencing*. Amsterdam: John Benjamins. doi: 10.1075/pbns.113
- Panther, K.-U., and Thornburg, L. L. (2004). The role of conceptual metonymy in meaning construction. *Metaphorik.de* 91–113.
- Panther, K. U., Thornburg, L. L., and Barcelona, A. (eds). (2009). *Metonymy and Metaphor in Grammar*. Amsterdam: John Benjamins. doi: 10.1075/hcp.25
- Peirce, C. S. (1960). *Collected Papers of Charles Sanders Peirce: Vol. I: Principles of Philosophy, and Vol. II: Elements of Logic*, eds C. Hartshorne and P. Weiss (Cambridge, MA: Harvard University Press).
- Peirsman, Y., and Geeraerts, D. (2006). Metonymy as a prototypical category. *Cogn. Linguist.* 17, 269–316. doi: 10.1515/COG.2006.007
- Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol.* 1:227. doi: 10.3389/fpsyg.2010.00227
- Radden, G. (2000). “How metonymic are metaphors,” in *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*, ed. A. Barcelona (Berlin: Mouton de Gruyter), 93–108.
- Ruiz de Mendoza Ibanez, F. J. (2017). “Metaphor and other cognitive operations in interaction: from basicity to complexity,” in *Metaphor: Embodied Cognition and Discourse*, ed. B. Hampe (Cambridge: Cambridge University Press), 138–159. doi: 10.1017/9781108182324.009
- Schank, R. C., and Abelson, R. (1977). *Scripts, Goals, Plans, and Understanding*. Hillsdale, NJ: Lawrence Erlbaum.
- Slobin, D. (1985). “Crosslinguistic evidence for the language-making capacity,” in *A Crosslinguistic Study of Language Acquisition. Theoretical Issues*, Vol. 2, ed. D. Slobin (Hillsdale, NJ: Lawrence Erlbaum), 157–1256.
- Sonesson, G. (2007). “The extensions of man revisited: from primary to tertiary embodiment,” in *Embodiment in Cognition and Culture*, eds J. M. Krois, M. Rosengren, A. Steidele, and D. Westerkamp (Amsterdam: Benjamins), 27–53.
- Streeck, J. (2009). *Gesturecraft: The Manu-Facture of Meaning*. Amsterdam: John Benjamins. doi: 10.1075/gs.2
- Streeck, J., Goodwin, C., and LeBaron, C. (eds). (2011). *Embodied Interaction: Language and the Body in the Material World*. Cambridge: Cambridge University Press.
- Sweetser, E. (2012). “Introduction Viewpoint and perspective in language and gesture,” in *Viewpoint in Language: A Multimodal Perspective*, eds B. Dancygier and E. Sweetser (Cambridge: Cambridge University Press), 1–22.
- Sweetser, E. E. (1990). *From Etymology to Pragmatics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620904
- Sweetser, E. E. (2007). “Looking at space to study mental spaces: co-speech gesture as a crucial data source in cognitive linguistics,” in *Methods in Cognitive Linguistics*, eds M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, and M. Spivey (Amsterdam: John Benjamins), 202–224. doi: 10.1075/hcp.18.15swe
- Sweetser, E. E. (2009). “What does it mean to compare language and gesture? Modalities and contrasts,” in *Crosslinguistic Approaches to the Psychology of Language: Studies in the Tradition of Dan Isaac Slobin*, eds J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, and S. Özçalışkan (New York, NY: Psychology Press), 357–366.
- Taub, S. (2001). *Language from the Body. Iconicity and Metaphor in American Sign Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511509629
- Waugh, L. R., and Monville-Burston, M. (1990). “Roman Jakobson: his life, work and influence,” in *Introduction to: Jakobson On Language*, eds L. R. Waugh and M. Monville-Burston (Cambridge, MA: Harvard University Press), 1–45.
- Wehling, E. (2017). Discourse management gestures. *Gesture* 16, 245–276. doi: 10.1075/gest.16.2.04weh
- Wilcox, P. (2004a). A cognitive key: metonymic and metaphorical mappings in ASL. *Cogn. Linguist.* 15, 197–222. doi: 10.1515/cogl.2004.008
- Wilcox, S. (2004b). Cognitive iconicity: conceptual spaces, meaning, and gesture in signed languages. *Cogn. Linguist.* 15, 119–148. doi: 10.1515/cogl.2004.005
- Wilcox, S. (2004c). Gesture and language: cross-linguistic and historical data from signed languages. *Gesture* 4, 43–73. doi: 10.1075/gest.4.1.04wil
- Wilcox, S., Wilcox, P., and Jarque, M. J. (2003). Mappings in conceptual space: metonymy, metaphor, and iconicity in two signed languages. *Jezikoslovje* 4, 139–156.
- Wolf, D., Rekkittke, L.-M., Mittelberg, I., Klasen, M., and Mathiak, K. (2017). Perceived conventionality in co-speech gestures involves the fronto-temporal language network. *Front. Hum. Neurosci.* 11:573. doi: 10.3389/fnhum.2017.00573
- Zima, E., and Bergs, A. (2017). Toward a multimodal construction grammar. Special issue. *Linguist. Vanguard* 3:s1. doi: 10.1515/lingvan-2016-1006
- Zlatev, J. (2014). Image schemas, mimetic schemas, and children’s gestures. *Cogn. Semiotics* 7, 3–29.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Mittelberg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Palm-Up Puzzle: Meanings and Origins of a Widespread Form in Gesture and Sign

Kensy Cooperrider^{1*}, Natasha Abner² and Susan Goldin-Meadow¹

¹ Department of Psychology, University of Chicago, Chicago, IL, United States, ² Department of Linguistics, University of Michigan, Ann Arbor, MI, United States

OPEN ACCESS

Edited by:

Marianne Gullberg,
Lund University, Sweden

Reviewed by:

Gaëlle Ferre,
University of Nantes, France
Maria Graziano,
Lund University, Sweden

*Correspondence:

Kensy Cooperrider
kensy@uchicago.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Communication

Received: 02 January 2018

Accepted: 22 May 2018

Published: 26 June 2018

Citation:

Cooperrider K, Abner N and
Goldin-Meadow S (2018) The
Palm-Up Puzzle: Meanings and
Origins of a Widespread Form in
Gesture and Sign.
Front. Commun. 3:23.
doi: 10.3389/fcomm.2018.00023

During communication, speakers commonly rotate their forearms so that their palms turn upward. Yet despite more than a century of observations of such palm-up gestures, their meanings and origins have proven difficult to pin down. We distinguish two gestures within the palm-up form family: the *palm-up presentational* and the *palm-up epistemic*. The latter is a term we introduce to refer to a variant of the palm-up that prototypically involves lateral separation of the hands. This gesture—our focus—is used in speaking communities around the world to express a recurring set of epistemic meanings, several of which seem quite distinct. More striking, a similar palm-up form is used to express the same set of meanings in many established sign languages and in emerging sign systems. Such observations present a two-part puzzle: the first part is how this set of seemingly distinct meanings for the palm-up epistemic are related, if indeed they are; the second is why the palm-up form is so widely used to express just this set of meanings. We propose a network connecting the different attested meanings of the palm-up epistemic, with a kernel meaning of *absence of knowledge*, and discuss how this proposal could be evaluated through additional developmental, corpus-based, and experimental research. We then assess two contrasting accounts of the connection between the palm-up form and this proposed meaning network, and consider implications for our understanding of the palm-up form family more generally. By addressing the palm-up puzzle, we aim, not only to illuminate a widespread form found in gesture and sign, but also to provide insights into fundamental questions about visual-bodily communication: where communicative forms come from, how they take on new meanings, and how they become integrated into language in signing communities.

Keywords: palm-up, gesture, sign, meaning, shrug, communication

INTRODUCTION

Centuries ago in Italy, a keen observer took notes on a conversation unfolding nearby. He inventoried a number of movements and gestures. He writes, for example: “Another with arms spread open showing the palm, shrugs his shoulders up to his ears and makes a grimace of astonishment” (qtd. in Isaacson, 2017, p. 283). The observer was Leonardo da Vinci, and his careful observations of the role of the body in everyday communication—including observations of a

deaf associate—informed his painting (Isaacson, 2017; see also Streeck, 2003). In *The Last Supper*, a panoramic buffet of gestural expression, Leonardo would capture the same gesture he described as “showing the palm.” The three apostles on the far right of the panel—Matthew, Thaddeus, and Simon—all produce some version of this form as they react to Jesus’s announcement that there is a traitor in their midst (**Figure 1**). The meaning of these palm-up gestures is impossible to pin down with precision, of course, but, in broad strokes, it is hard to mistake: the three men are taken aback and trying to make sense of what has just happened. *What? Who? Could it be?* Across half a millennium and despite peeling paint, the gestures still convey meaning.

Palm-up gestures remain ubiquitous today, and yet their meanings are still difficult to pin down precisely; they are generally interpretable and yet confoundingly elusive, much as in Leonardo’s rendering. The present paper shines a spotlight on this family of forms, exposing a puzzle of form and meaning that is bigger and thornier than previously appreciated. We are hardly the first to note the prominence of palm-ups in communication. Indeed, observations about the form stretch back to antiquity, and a growing number have been made in recent decades. Unfortunately, these observations have sometimes been made in mutual isolation from each other, often using different analytic frameworks and pursuing different ends. Work on palm-ups as used by speakers, for instance, has often been carried out independently from work on palm-ups as used by signers. An important goal of the present paper is to bridge these literatures and present a comprehensive synthesis of research on palm-ups in both spoken and signed communication. Another goal looms perhaps larger: shedding light on palm-up forms stands to shed light also on general questions about where bodily communicative forms come from, how they take on new meanings through processes of semantic and pragmatic

extension, and how they become integrated into language in signing communities. That is, palm-ups—in all their ubiquity and multiplicity of meanings—present a critical case study for scholars of *visual-bodily communication*.

Part of what makes palm-ups a compelling phenomenon of study is their sheer pervasiveness. The largest study of the gestures accompanying spoken English to date, an analysis of 8000 gestures produced by 129 speakers, found that two gesture types involving the palm-up form together accounted for 24% of all gestures (Chu et al., 2013). Such ubiquity has also been reported in analyses of signed communication. A large-scale corpus study of British Sign Language (BSL) reports that the palm-up, glossed by the authors as a discourse marker meaning WELL, is the second most frequent sign (after the first-person pronoun) (Fenlon et al., 2014). A comparably sized corpus study also found the palm-up to be the second most frequent sign in the sign language of Australia (Auslan) (Johnston, 2012), as did a smaller corpus study of New Zealand Sign Language (NZSL) (McKee and Wallingford, 2011); and it was the third most frequent sign in a corpus study of Swedish Sign Language (SSL) (Börstell et al., 2016).

There is little reason to think the prominence of palm-up forms is a quirk of communication in Anglophone and European communities. As we review later, speakers from communities in Asia, Africa, South America, and elsewhere also produce this same form family—and one “family member” in particular—to express a number of meanings. Several of these meanings seem at first blush quite distinct from each other—sometimes even contradictory—and yet, intriguingly, the same cluster of meanings pops up in culture after culture. Moreover, signers from far-flung signing communities produce the palm-up form to express a similar cluster of meanings. Palm-ups are also used in emerging languages, including the so-called “shared” sign



FIGURE 1 | Detail from Leonardo da Vinci's *The Last Supper*, a mural from the late fifteenth century, showing three figures producing gestures with the palms turned up.

systems of villages with high rates of hereditary deafness (Nyst, 2012) and the idiosyncratic communication systems innovated by profoundly deaf people who grow up without access to conventional sign language, called *homesigners* (e.g., Franklin et al., 2011). These observations invite two preliminary points. First, if a community uses a single form to cover two or more meanings—and, if other communities also use a single form to cover the same meanings—these meanings are most likely related to each other. The task of the analyst—and the task we take up here—is to articulate these links. Second, if different communities use the very *same* form for a particular set of meanings, there is probably a motivated relationship between form and meaning, however obscure this motivation may be to the casual observer (for arguments of this type in sign, see Aronoff et al., 2005; Wilbur, 2010). The challenge for the analyst—and, again, the challenge we take up here—is to try to discern that motivation.

Our attempt to understand palm-ups thus engages long-standing questions about form and meaning common to the study of all communication—spoken, signed, or gestured. The “palm-up puzzle” of our title can be crystallized as having two parts. First, how are the seemingly distinct *meanings* of palm-ups related? And, second, why is the palm-up *form* used to express them? Adding to the difficulty of our task, there are complexities of interpretation that make the puzzle thornier than it first appears. Most critically, there is not one palm-up form but a family of forms, and whether this form family has important finer divisions within it remains contested. Moreover, palm-ups often co-occur with other bodily forms, including facial expressions, head shakes, and, notably, shoulder shrugs, as we discuss in detail later. Terminological profusion further complicates matters. Gestures exhibiting the palm-up form have been called “hand shrugs” (Johnson et al., 1975), “palm-revealing” gestures (Chu et al., 2013), “hand flips” (Ferré, 2011), “palm up open hand” (PUOH) gestures (Müller, 2004), the “open hand supine” gesture family (Kendon, 2004), and the “rotated palms” gesture family (Gawne, 2018) (see also Givens, 2016, p. 1–2). Gesture researchers have variously classified palm-ups as “emblems” (e.g., Johnson et al., 1975), “recurrent gestures” (Müller, 2017), “pragmatic gestures” (Kendon, 2004), “interactive gestures” (Bavelas et al., 1992), “metaphorics” (e.g., Parrill, 2008), and beyond. Sign researchers have described palm-ups as having both grammaticalized uses as a sign and uses as a co-sign gesture (McKee and Wallingford, 2011). When used as a sign, grammatical classification varies; some describe palm-up signs as discourse markers (Engberg-Pedersen, 2002), others as particles (Conlin et al., 2003). Palm-up forms have been described as serving a suite of functions—cohesive, modal, and interactive—all within individual sign languages (Engberg-Pedersen, 2002; McKee and Wallingford, 2011; Volk, 2017). There are also clear cases where frozen signs, such as interrogative markers and indefinites, incorporate the palm-up form (e.g., Zeshan, 2004). In short, in both gesture and sign, palm-ups exhibit wide diversity in form, puzzling multiplicity in meaning, and vexing variability in the terminology and frameworks used to characterize them.

Our plan for taking on the palm-up puzzle is as follows. We begin by collating existing observations about how palm-up forms are used by gesturers (next section) and by signers

(following section). As will become clear, our focus is not on the entire family of palm-up forms, but on a particular gesture within the family that is widely used in both gesture and sign: what we call the *palm-up epistemic gesture*. This particular palm-up presents a puzzle all on its own, and we consider it in detail: we discuss the meanings that have been most widely associated with this gesture across speaking and signing communities, and we propose a meaning network to account for how these meanings are related to each other. Recognizing the provisional status of this proposal, we also discuss the types of new evidence that would be most useful in testing and refining it. Finally, before concluding, we evaluate two accounts of the ultimate origins of the palm-up epistemic gesture—that is, accounts of what motivates the use of this form for this set of meanings. Ultimately, by treating this particular palm-up gesture in detail, we aim to shed light on the meanings and origins of the entire palm-up form family.

PALM-UPS IN GESTURE

Preliminaries

Before cataloging observations about use of the palm-up form in gesture, it will be helpful to briefly introduce our perspective on gestural form and meaning. Here, we use the term “form” to refer to features of handshape and movement irrespective of meaning, and reserve “gesture” for particular conventional pairings of form and meaning (for discussion, see Cooperrider and Núñez, 2012). Take the “thumbs up”: speakers may produce a number of gestures involving a thumbs-up *form*, but the “thumbs-up gesture” refers to a conventional pairing of that form with a particular meaning. The form/gesture distinction is even more critical in light of the fact that conventional gestures, like language, sometimes exhibit cases of apparent “homonymy” in which the same form is associated with distinct meanings (for discussion, see Sherzer, 1973). As an illustration, consider two conventional gestures involving an extended index finger. On one occasion, a speaker may hold up this finger vertically to raise an objection; on another, the vertical finger may serve to ask an interlocutor to wait a second. These uses share a common *form*, but they have different meanings and motivations. In short, they are different *gestures*.

With this perspective in mind, we now turn to several sticking points in the palm-up literature. The first and biggest of these is the question of whether all forms that exhibit an upturned palm should be considered one inter-connected family of gestures, a family but with key divisions, or perhaps several distinct gestures related in form only. Researchers broadly agree that the family of palm-up forms is sprawling; they disagree on how the family should be divided or whether it should be. Kendon (2004, pp. 273–281), for instance, divides the family based on differences in motion pattern. One variant—the “palm lateral”—involves rotating the forearms so that the palms face upward and moving the hands apart. A second—the “palm presentation”—involves moving the upturned palm toward the listener, as if “presenting” something. Other authors have echoed such divisions. Chu et al. (2013), for example, distinguish between “palm-revealing gestures” (comparable to Kendon’s “palm lateral”) and “conduit

gestures” (comparable to his “palm presentation”). In motivating such a split, these researchers argue that formational features can generally (though not always) distinguish between these two variants, and that these variants express different kinds of meanings.

Other authors make no such divisions, seeing the palm-up as a large extended gesture family, a set of related forms paired with related meanings. For instance, Müller (2004) considers these formational variants together under a unified semantic theme, as we discuss later. Streeck (2009) would seem to similarly lump together uses, describing some uses that seem akin to Kendon’s “palm presentation” and others akin to the “palm lateral.” Other researchers follow this lumping approach while also taking the presentational variant of the gesture as the central explanandum (e.g., McNeill, 1992; Parrill, 2008). In the sign literature, to which we turn later, Engberg-Pedersen (2002) echoes this focus on the presentational use, as do other researchers to lesser degrees (e.g., McKee and Wallingford, 2011). Part of what makes this sticking point particularly sticky is that researchers are not always explicit about how, if at all, they divide up the palm-up form family.

Our own strategy here will be a splitting one. Similar to the proposals of Kendon (2004) and Chu et al. (2013), we draw a key distinction within this extended form family between what we term “palm-up epistemic”¹ gestures (comparable to Kendon’s “palm lateral” and Chu et al.’s “palm-revealing”) and “palm-up presentational” gestures (comparable to Kendon’s “palm-presentation” and Chu et al.’s “conduit gestures”) (Figure 2). There are certainly uses of palm-up forms that do not fall into either category, but we narrow our analysis to these². For clarity, we label these gestures with reference to the most salient aspect of their form (“palm-up” in both cases) and the most salient aspect of their meaning (“epistemic” or “presentational”). In their prototypical versions in Anglophone and European cultures, these gestures appear to exhibit differences in form: in the prototypical palm-up epistemic gesture, the palm or palms are turned upward and moved outward; in the prototypical palm-up presentational, the palm is directed toward the interlocutor³. However, we caution that form is not a foolproof guide to meaning (see also McKee and Wallingford, 2011; Chu et al., 2013). In many tokens of palm-up gestures, interlocutors—and analysts—will need to assign meaning based on context in addition to form, and some tokens may be compatible with either an epistemic or presentational meaning.



FIGURE 2 | Examples of two gestures within the palm-up form family. In English speakers, depicted here, these gestures prototypically involve different motion patterns. Palm-up epistemic gestures (**Left**) involve a lateral separation of the hands (or a lateral movement of one hand), and are used to express epistemic meanings. Palm-up presentational gestures (**Right**) prototypically involve a movement toward the interlocutor as if “presenting” an idea. Images reproduced under fair use.

Again, while we aim to shed light on the entire palm-up form family, our particular focus in much of what follows is the epistemic variant. This is for a few reasons. Both types of palm-up gestures are common in spoken communication (e.g., Chu et al., 2013) and both appear to be used in signed communication (e.g., Engberg-Pedersen, 2002), yet the palm-up epistemic appears to be much more widely incorporated into sign language grammars (e.g., as question-markers or modals). A likely reason for this difference is that the palm-up epistemic gesture has several highly conventional, readily glossable uses (e.g., “I don’t know”); it may thus be considered a “holophrastic emblem” in that it can serve as a communicative turn all on its own⁴. Much research suggests that highly conventional gestures, including holophrastic emblems, are ripe for incorporation into sign systems (Wilcox, 2004; Pfau and Steinbach, 2006; Spaepen et al., 2013; Haviland, 2015). The palm-up presentational gesture, by contrast, is used to underline the presentational function of speech rather than replace speech, and does not appear to have an easily glossable, holophrastic meaning. Another reason for our focus on the palm-up epistemic is that it is perhaps the more puzzling of the two gestures. Though the palm-up presentational gesture is not as easily glossed, its meaning appears to be consistent across uses—it underlines the presentational aspect of speech. The palm-up epistemic, however, is widely

¹We use “epistemic” in the general sense of “relating to knowledge,” not in any of the technical senses in which it is sometimes used in linguistics or philosophy.

²We do not, for example, discuss Kendon’s category of “palm-addressed” gestures (Kendon, 2004). Nor do we consider the palm-up, open-hand points sometimes used by speakers when indicating interlocutors or other people (see, e.g., McGowan, 2010 for examples), perhaps for politeness reasons (Calbris, 1990). Finally, we do not discuss the emblematic “all gone” gesture, produced with both palms turned up, which is widely observed in children’s gesture and in child-directed gesture (McNeill, 1992; Iverson et al., 2008; Beaupoil-Hourdel and Debras, 2017), but does not appear to be widely used in adult discourse.

³Though not prototypical, palm-up epistemic gestures also sometimes involve an indicating component, as when the speaker asks a question about some person or entity in the world. Matthew’s gesture in *The Last Supper* (see Figure 1, at left), apparently directed toward the center of the panel, can be interpreted in this way—as reinforcing a question about Jesus.

⁴This high degree of conventionalization is likely the same reason that the palm-up epistemic gesture is enshrined in GIFS (see, e.g., <https://giphy.com/search/shrug>) and emoji (see, e.g., <https://emojipedia.org/shrug/>). And it may be the same reason that palm-up epistemics are produced earlier in child development than palm-up presentation gestures (Graziano, 2014).

associated with a seemingly disparate set of meanings beyond its most conventional ones.

Our division between palm-up epistemic gestures and palm-up presentational gestures makes all the more sense in light of a second sticking point in the literature: the shrug. Though this fact sometimes goes unmentioned, there is a clear affinity between the palm-up epistemic gesture and shoulder shrugs—indeed, the two commonly co-occur, and some have even considered them functionally interchangeable (e.g., Chu et al., 2013). Palm-up presentational gestures, meanwhile, have no such affinity with the shrug. Research on the shrug is scarcer than research on palm-ups, but detailed observations go back to Darwin (1998/1872), who considered the shrug a “gesture natural to mankind” (p. 269), and there are also a handful of more recent studies (Givens, 1977, 1986; Debras, 2017; Jehoul et al., 2017). As we will argue later, understanding the relationship between the shrug and the palm-up epistemic gesture may be critical to understanding the broader palm-up puzzle.

A third and final sticking point concerns what kind of gestures palm-ups are. If researchers agree on anything, it is that palm-ups are interactional in nature—that is, they are not like pointing or depicting gestures that relate to the content of what is being described. Some thus describe them as “interactive” (Bavelas et al., 1992), “speech-handling” (Streeck, 2009), or “pragmatic” (Kendon, 2004) gestures. Some researchers also consider them fundamentally metaphoric, as gestural expressions of the “conduit metaphor” (McNeill, 1992). Another point of disagreement concerns whether palm-ups should be considered conventional “emblems,” “recurrent gestures” (see Müller, 2017), or more idiosyncratic in nature. Evidently, they run the whole gamut: like emblems, they sometimes have a readily glossable meaning (Johnson et al., 1975), but like “recurrent” or idiosyncratic gestures, they can also be layered atop utterances to add shadings of meaning (Gawne, 2018). With these sticking points in mind, we can now move to an overview of observations about palm-up forms, with a focus on the meanings that have been ascribed to the palm-up epistemic gesture.

Discussions of Palm-Ups in Gesture

Observations about palm-ups have a long history (reviewed in Müller, 2004). Quintillian mentions them briefly in his classic discussion of gesture in oratory; John Bulwer notes how a palm-up form is used when “begging;” and Andrea de Jorio describes several contexts in which Neapolitan speakers make use of the form family. The first detailed treatments, however, can be found in Kendon (2004) and Müller (2004). Kendon’s (2004) discussion characterizes how the gesture is used by English speakers in Britain and Italian speakers in Naples. As noted earlier, a key feature of his treatment is the separation of the larger family of palm-up forms into three distinct variants: the “palm lateral” (in our terms, the “palm-up epistemic”), the “palm presentation” (in our terms, the “palm-up presentational”), and the “palm addressed” (which we do not discuss). Our focus, again, is on the first of these, the palm lateral, which corresponds most closely to what we term the palm-up epistemic gesture. Kendon distinguishes five uses of this variant: (1) unwillingness or inability on the part of the speaker; (2) that a proposition

is obvious; (3) as part of a question that cannot or need not be answered (i.e., a rhetorical question); (4) that a proposition “could be;” (5) the availability of the speaker for service. He discerns an over-arching theme of “non-intervention” running through these uses⁵. Interestingly, Kendon does comment on the apparent affinity between certain of these uses and the shoulder shrug (p. 265), but he does not dwell on the question of what might explain this affinity.

Müller (2004) presents a rich history of observations about palm-ups—or, in her terms “palm-up open hand” (PUOH) gestures—as well as examples of their use in Spanish speakers. She sees the PUOH as part of an extended family of gestures, but, unlike Kendon, she does not tease out major subdivisions of the family based on motion pattern. For Müller, this family is rooted in practical actions of *giving* and *receiving* objects. Thus, she sees PUOH gestures as fundamentally metaphorical in that they treat the abstract objects of discourse—propositions, ideas, questions, answers—like the physical objects of everyday life in that they can be held up, offered, requested, exchanged, and so on.

The giving-related senses Müller identifies for the PUOH center around the idea that there is an imaginary object presented on the palm. These senses include: (1) presenting an abstract object as visible or even obvious; (2) presenting an abstract object for joint inspection; (3) proposing a shared perspective on an abstract object. The receiving-related senses Müller identifies center around the idea that the palm is shown to be empty. These include: (1) to plead for an abstract object; (2) to request an abstract object; (3) to express openness to the reception of some abstract object; (4) to express the fact of not knowing. As noted, Müller does not highlight the contrasts in motion pattern that Kendon (2004) keys in on, nor does she dwell on the affinity between some palm-ups and the shrug. Müller’s treatment thus combines uses of the palm-up that Kendon treats as distinct and gives them a single overarching motivation. In some cases, this leads the two authors to different conclusions about what motivates the use of this form. For example, Müller relates the use of the PUOH to express “obviousness” to the idea that some imaginary object is presented forcefully and thus “obviously” on the palm, whereas Kendon relates this use to the idea of non-intervention, that “nothing further can be said” (p. 265).

A more recent analysis by Streeck (2009) combines elements of Kendon’s and Müller’s accounts. Like Kendon, for instance, he notes the affinity of meaning between certain uses of the palm-up and the shrug. Like Müller, he emphasizes the grounding of palm-up gestures in practical actions. His treatment is also more explicitly steeped in the idea that palm-ups embody metaphors, in particular the “conduit metaphor” (Reddy, 1979; McNeill, 1992), according to which conversation is conceptualized as the exchange of abstract objects through a channel, or “conduit.” A distinctive aspect of Streeck’s treatment is his attention to use of palm-ups in certain types of interactive moves, also a focus of

⁵This fifth use is given the least attention in Kendon’s account, and it is the one that conspicuously does not fit the theme of “non-intervention” that he discerns. A different interpretation is that this use stems from the act of moving the hands away from the body and presenting the torso, as if to say, “Here I am,” which may be yet another conventional gesture in the palm-up form family.

some sign language analyses (e.g., Engberg-Pedersen, 2002). For instance, one of the uses he discusses is the “weak offering” in which the palm-up suggests that some idea is being offered up, but without particular conviction or assertiveness.

The only large-scale quantitative analysis of palm-up gestures to date comes from a study by Chu et al. (2013) on individual differences in gesture production. The authors separate out “conduit gestures” (corresponding to Kendon’s “palm presentation” variant and our “palm-up presentational” gesture) from what they term “palm-revealing gestures” (corresponding to Kendon’s “palm lateral” and our “palm-up epistemic” gesture). They ascribe three primary uses to the palm-revealing variant: to express uncertainty, to express resignation, or to show that the speaker has nothing more to say. Interestingly, they also note a possible motivation for the link between the palm-up form and the meanings they ascribe to it: that the hand is shown to be “empty.” Thus, like Streeck, they echo Müller’s interpretation that the palm-up embodies metaphors of giving and receiving. Interestingly, more than other authors to date, Chu et al. (2013) emphasize the affinity of palm-revealing gestures with the shrug. In their coding scheme, for example, shoulder shrugs produced without a palm-up were considered palm-revealing gestures if “used for the same purposes” (p. 700). Finally, they are notably cautious about the possibility of distinguishing these two types of palm-up gestures on the basis of form alone. Though Kendon’s description of the “palm lateral” suggests outward movement is always present, Chu et al. note that their “palm-revealing” gestures do not always “move laterally and the palm may not always face upward” (p. 700).

Finally, we turn to the most thorough analysis of a palm-up gesture in a non-European language—Syuba, a Tibeto-Burman language of Nepal (Gawne, 2018). Gawne discusses in particular the “rotated palms gesture family,” which in Syuba is associated with a theme of interrogativity. One of the most interesting aspects of this palm-up gesture—which we take to be a version of the palm-up epistemic gesture—as it is used by Syuba speakers is that it involves a distinctive handshape not reported elsewhere: the index finger and thumb are extended and the remaining fingers curled back into the palm to various degrees. Moreover, the Syuba version does not prototypically involve a lateral movement of the hands, or any other distinctive motion pattern. In terms of its meaning, the emblematic form of the gesture is produced without speech to ask, “What are you doing?” “What to do?” or “What to say?” Such emblematic uses, Gawne notes, are widely attested across India and Nepal and are likely related to formationally similar interrogative signs in the sign languages of the region (see next section). Gawne also observes substantial variation in the gesture’s form—it may involve one or two hands, more or less curling-in of the fingers, and a substantial hold or no hold at all. To the gesture in its co-speech uses, Gawne ascribes meanings of *interrogativity*, *uncertainty*, and—intriguingly— *hypotheticality*. Finally, Gawne notes that the gesture is sometimes accompanied by a shrug, particularly in its emblematic meaning of “What to do?”

A number of other observations have been made across cultures about what appear to be uses of the palm-up epistemic gesture. Many are in-passing comments, but it is

nonetheless striking that most of the uses of the gesture just discussed have also been described in other, often unrelated communities. Discerning such commonalities in meaning involves interpretation on our part, as researchers do not always use the same descriptors for different uses of the gesture. This caveat aside, we group these observed uses into six meaning categories: *absence of knowledge, ability, or concern*; *uncertainty*; *interrogatives*; *hypotheticality*; *obviousness*; and *exclamatives* (Table 1; see also Appendix 1 in Supplementary Material for a reorganization of the same data, along with additional details from primary sources).

As suggestive as the evidence is about the pervasiveness and recurring meanings of the palm-up epistemic, it also has limitations. For one, many of these treatments do not include

TABLE 1 | Observed uses of the palm-up epistemic in gesture.

Meaning category	Communities
Absence of knowledge, concern, or ability <i>Example glosses:</i> “I don’t know” “What can I do?”	Arabic (Barakat, 1973) Catalan (Payrató, 1993) English (American) (Ekman and Friesen, 1972; Johnson et al., 1975; Bavelas et al., 1992; Streeck, 2009) English (British) (Kendon, 2004; Chu et al., 2013) Farsi (Iran) (Sparhawk, 1978) Italian (Neapolitan) (Kendon, 2004) Kenya (Creider, 1977) Russian (Monahan, 1983) Spanish (Colombia) (Saitz and Cervenka, 1972) Spanish (Spain) (Müller, 2004) Syuba (Nepal) (Gawne, 2018) Yoruba (Nigeria) (Agwuele, 2014) Zulu (South Africa) (Brookes, 2004)
Uncertainty <i>Example glosses:</i> “I’m not sure” “Maybe”	Arabic (Barakat, 1973) Catalan (Payrató, 1993) English (American) (Johnson et al., 1975) English (British) (Chu et al., 2013) French (Ferré, 2011) Syuba (Nepal) (Gawne, 2018)
Interrogatives <i>Example glosses:</i> “What?” “Where?” “Who?” “Are you coming?”	English (American) (McNeill, 1985) English (British) (Chu et al., 2013) Italian (Neapolitan) (Kendon, 2004) Kenya (Creider, 1977) Portuguese (Brazilian) (Rector, 1986) Spanish (Colombia) (Saitz and Cervenka, 1972) Spanish (Spain) (Müller, 2004) Syuba (Nepal) (Gawne, 2018)
Hypotheticals <i>Example glosses:</i> “I hope he will come.”	Syuba (Nepal) (Gawne, 2018)
Obviousness <i>Example glosses:</i> “Of course he did”	Dutch (Jehoul et al., 2017) English (British) (Kendon, 2004) French (Calbris, 1990) Italian (Neapolitan) (Kendon, 2004)
Exclamatives <i>Example glosses:</i> “Whatever!”	Portuguese (Brazilian) (Rector, 1986)

fine-grained descriptions of form, so we cannot be sure that the prototypical motion pattern of the gesture described earlier is found more broadly—in at least one case, a palm-up gesture with epistemic meanings uses a different prototypical form (Gawne, 2018). Further, given that many of these sources do not attempt an exhaustive treatment of the gesture, the lack of mention of any particular meaning should not be taken as evidence that the meaning is absent from a community. A final limitation of the literature is that, even among the extended treatments of the palm-up, quantitative methods are rare (but see, e.g., Chu et al., 2013; Jehoul et al., 2017). Nevertheless, the widespread use and apparent semantic regularities in the palm-up epistemic are striking. A natural further question is whether the palm-up form is *universally* used to express these meanings—that is, is the palm-up epistemic *gesture* found in all communities? Any attempt to answer this question would be premature, and absolute universals are notoriously difficult to demonstrate. What we can say, however, is that use of the gesture to express a recurring set of meanings strongly suggests (a) that the meanings are related and (b) that the use of the palm-up form to express them is motivated. We revisit the puzzle presented by these observations later, after first considering comparable evidence from the palm-up in sign.

PALM-UPS IN SIGN

Preliminaries

The palm-up form in sign languages has also been widely studied. Though this line of research often nods to possible relations between the gestural and signed uses of the form (see, e.g., Zeshan, 2004, p. 23; Van Loon et al., 2014), little work has directly engaged with both sizeable literatures at once. Several of the sticking points bedeviling work on palm-up forms in gesture are evident here, too—for instance, whether there is more than one form-meaning pairing at work, whether palm-ups are related to the shrug in some way, and how palm-ups should be classified. This last sticking point takes on new significance in the sign literature because one's choice of terminology is bound up with fraught empirical and theoretical issues. Whether palm-ups are considered lexical items, discourse markers, or co-sign gestures has implications—not only for the analysis of this particular form—but for general questions about differences between sign and gesture (e.g., Goldin-Meadow and Brentari, 2017), and about how sign languages may draw on gestures from surrounding speaking communities. In what follows, we focus on the most extended and focused discussions of palm-ups in sign; we begin with palm-ups in well-established sign languages and then turn to homesign systems.

Discussions of Palm-Ups in Sign

One of the earliest in-depth treatments of the palm-up in sign is Conlin et al.'s (2003) analysis of the form in American Sign Language (ASL). The authors note that the form has clear lexical incarnations—such as in the signs WHAT and MAYBE—as well as uses as a discourse particle indicating uncertainty of different kinds. They focus, in particular, on a use of the particle to mark “indefiniteness.” Though difficult to lexically gloss, the addition

of the palm-up can turn a sign sentence that means “A boat sank off Cape Cod” into a sentence with a more indefinite meaning, such as “*Some boat or other* sank off Cape Cod” or “*Some kind of* boat sank off Cape Cod” (p. 8). Depending on its position, the palm-up may also express indefiniteness, or uncertainty about the proposition as a whole, e.g., “A boat sank off Cape Cod *I think*” (p. 10). Such uses of the palm-up thus allow one to sidestep conversational norms limiting contributions to those known to be true (Grice, 1975). Conlin and colleagues also briefly describe uses of the palm-up form for emphasis, as in “John does not know the answer!” (p. 13), noting that such an utterance implicitly asks the question, “How could you have thought he would?” And, finally, they characterize several uses of the form in sentences with HOPE and WISH. They link these uses to the broader theme of uncertainty; but, as discussed later, we group these with hypotheticals and other statements of possibility.

Aboh et al. (2005) also pursue a particle analysis for a palm-up sign glossed as G-WH (general WH-word) in Indian Sign Language (IndSL, also called Indo-Pakistani Sign Language; see Zeshan, 2003, where the sign is labeled as KYA:). As in the “rotated palms” gesture in Syuba, the palm-up particle in IndSL has a notable handshape, with the index finger and thumb extended and the other fingers curled slightly into the palm. This sign form can also be used as a sentence-final discourse particle signaling hesitation and as an indefinite marker. The interrogative and indefinite uses of the form also share other syntactic characteristics. Though G-WH is the only specifically interrogative sign in IndSL, it combines with other signs to express more specific interrogative meanings. For example, FACE G-WH can be used to ask “Who?” (Aboh et al., 2005). Similarly, the palm-up form can be combined with the sign MAN to form the indefinite SOMEONE/SOME MAN (Zeshan, 2003). This use of a specific word or morpheme to form paradigms of indefinite and WH-expressions is common across spoken languages.

Around the same time, Engberg-Pedersen (2002) analyzed the palm-up form in Danish Sign Language (Dansk Tegnsprog, DTS). She describes it as fundamentally “presentational,” a “materialization of the conduit metaphor” (p. 143). Like Conlin et al. (2003), she notes that the form appears to be present in certain clearly lexical DTS signs, such as WHAT and WHERE. Her focus, however, is on uses of the form as a “gesture;” she primarily focuses on analyzing how the form is placed in interactive sequences, rather than on identifying its invariant meanings. Though this is not her goal, many of the uses she illustrates bear a clear relation to those identified in ASL and IndSL. These include cases where the signer is expressing uncertainty or tentativeness, or is asking a question. However, it should also be noted that she describes a number of other uses for the palm-up that are hard to square with the observations of other sign researchers. A possible reason for this discrepancy is that Engberg-Pedersen explicitly treats presentational and epistemic uses of the palm-up under the same umbrella, in the same way that some gesture researchers do (e.g., Müller, 2004).

More recently, McKee and Wallingford (2011) have analyzed the palm-up in NZSL, characterizing it as a “frequent and multi-functional item” (p. 240). They are explicit about the thorniness of classifying this “item,” alternately describing it as a “gesture,”

a “sign,” or simply a “form.” They note, too, that the palm-up form exhibits wide formational variation; and they report NZSL signers’ intuitions that such “variations are neither consistent in usage, nor necessarily contrastive in meaning” (p. 220). Their data consist of a corpus of conversational signing, produced by 20 signers, totaling more than 5,000 signs. They find that the palm-up form accounts for 5% of all signs, making it the second most frequent sign in their corpus [comparable to what has been reported for BSL (Fenlon et al., 2014) and Auslan (Johnston, 2012)]. Following Engberg-Pedersen (2002), the authors focus on the sequential positioning and functioning of the form rather than on its invariant meanings. However, they too note a number of uses for the form that align with those epistemic uses reported elsewhere, including: expressions of uncertainty, interrogatives, hypotheticals, expressions of obviousness, and exclamatives. Intriguingly, they also note an “elaborative” use of the form, in circumstances that resemble the use of English “which” to introduce free relative clauses (p. 229).

Two especially valuable sets of observations about epistemic uses of the palm-up come from studies of *homesigners*, profoundly deaf individuals raised without access to a conventional language (e.g., Goldin-Meadow and Mylander, 1984; Goldin-Meadow, 2003). The first of these is a study of an adult homesigner (and her hearing associates) from the Enga region of Papua New Guinea (Kendon, 1980). Kendon describes a number of uses of what he calls “the double palm present” and its cousin the “lateral hand flip,” which have related meanings. These are used as a question marker [example utterance: “Whose father is coming?” (p. 276)]; to indicate the absence of knowledge [“I don’t know” (p. 277)]; and, less commonly, in a “whether” statement [“Whether he will... that’s his business” (p. 276)], which is similar in meaning to uses of the palm-up for hypotheticals reported elsewhere. Kendon discerns a theme running through these uses—“absence of knowledge” (p. 278)—and notes in passing that the double palm present is likely related to the shrug. Finally, he also observes that the form is used in certain contexts to express negation.

The second set of observations focused on a child homesigner in the United States (Franklin et al., 2011). The researchers analyzed 208 uses of the “flip,” as they call the palm-up epistemic, which were observed in a corpus of 3,080 gestured utterances that the signer, David, produced between the ages of two and four. Three primary uses of the gesture were observed. A first was to mark questions [e.g., “Why is the car there?” (p. 8)]—indeed, 92% of questions in the dataset involved a flip, while others were marked by a facial expression or, interestingly, a shrug. Of comparable frequency was the use of the flip as an exclamation—that is, to mark heightened affect. Examples included cases in which David was showing frustration (“Whatever!”) or expressing surprise. Under the umbrella of exclamative use, the researchers also included expressions of “doubt,” a use of the palm-up epistemic observed in hearing gesturers and already discussed. Finally, a rarer but intriguing use of the form turned up in David’s expressions about location. In one example, David combined a palm-up form with a pointing sign to create an utterance glossed as “The place where the puzzle goes is the toy bag” (p. 407). The authors interpret this use as analogous to what

are sometimes called “free relative” expressions. In fact, as the authors observe, these three uses—interrogatives, exclamatives, and relatives—are tacitly connected in English and other spoken languages through their common use of interrogative words.

A number of further observations have been made about what appears to be the palm-up epistemic in other signing communities (Table 2; see Appendix 2 Supplementary Material for a reorganization of the same data, with additional details), though often in passing. Most interesting for our purposes, these observations, taken together, touch on all of the meaning categories ascribed to the palm-up epistemic in co-speech gesture and discussed earlier: expressions of absence of knowledge,

TABLE 2 | Observed uses of the palm-up epistemic in sign.

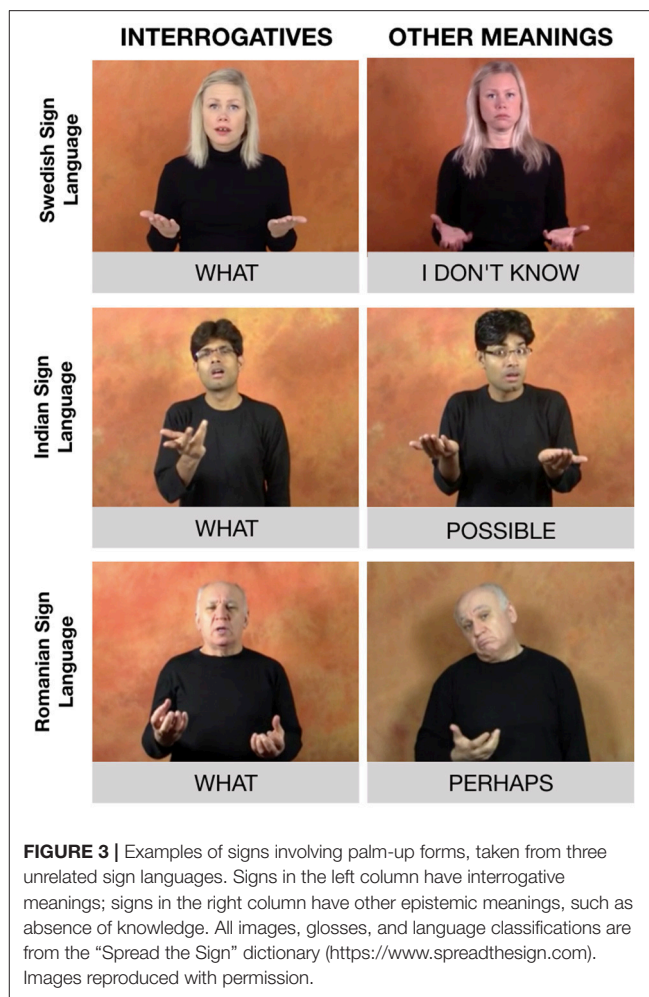
Meaning category	Communities
Absence of knowledge, concern, or ability <i>Example glosses:</i> “I don’t know” “What can I do?”	Enga homesign (Papua New Guinea) (Kendon, 1980)
Uncertainty <i>Example glosses:</i> “I’m not sure” “Maybe”	American Sign Language (Conlin et al., 2003) Danish Sign Language (Engberg-Pedersen, 2002) Indo-Pakistani Sign Language (Zeshan, 2003) Inuit Sign Language (Schuit, 2014) New Zealand Sign Language (McKee and Wallingford, 2011) Sign Language of the Netherlands (Van Loon, 2012) Turkish Sign Language (Zeshan, 2006b) US homesign (Franklin et al., 2011)
Interrogatives <i>Example glosses:</i> “What?” “Where?” “Who?” “Are you coming?”	American Sign Language (Conlin et al., 2003) Danish Sign Language (Engberg-Pedersen, 2002) Enga homesign (Papua New Guinea) (Kendon, 1980) Finnish Sign Language (Zeshan, 2004) Indo-Pakistani Sign Language (Zeshan, 2003) Inuit Sign Language (Schuit, 2014) Israeli Sign Language (Meir, 2004) New Zealand Sign Language (Zeshan, 2004) Providence Island Sign Language (Washabaugh et al., 1978) Sign Language of the Netherlands (Van Loon, 2012) Tanzania Sign Language (Zeshan, 2013) Turkish Sign Language (Zeshan, 2006b) US homesign (Franklin et al., 2011) Yucatec Maya Sign Language (Shuman and Cherry-Shuman, 1981) Yolu Sign Language (Bauer, 2015)
Hypotheticals <i>Example glosses:</i> “I hope he will come.”	American Sign Language (Conlin et al., 2003) Enga homesign (Papua New Guinea) (Kendon, 1980) New Zealand Sign Language (McKee and Wallingford, 2011)
Obviousness <i>Example glosses:</i> “Of course he did”	New Zealand Sign Language (McKee and Wallingford, 2011)
Exclamatives <i>Example glosses:</i> “Whatever!”	American Sign Language (Conlin et al., 2003) New Zealand Sign Language (McKee and Wallingford, 2011) US homesign (Franklin et al., 2011)

concern, or ability; expressions of uncertainty; interrogatives; hypotheticals; expressions of obviousness; and exclamatives (for examples, see **Figure 3**). Several other meanings ascribed to the palm-up epistemic in sign do not have a clear counterpart in the existing gesture literature, however. For example, the use of the palm-up for indefinites (*someone, somewhere, somehow*) has been described in both ASL (Conlin et al., 2003) and IndSL but not in any hearing community to date. These uses may be closely related to the interrogative uses of the palm-up. After all, though not the case in English, it is common cross-linguistically for indefinite expressions to be formed out of question words (Ullan, 1978; Haspelmath, 1997), as noted above in the discussion of G-WH in IndSL. Further, several authors also note that the palm-up is used to express negation in certain contexts, a phenomenon observed in Turkish Sign Language (Zeshan, 2006b), Inuit Sign Language (Schuit, 2014), and Enga homesign (Kendon, 1980). Finally, observations of the palm-up in ASL (Conlin et al., 2003) and in US homesign (Franklin et al., 2011) note uses of the palm-up in non-restrictive and free relative clauses—intriguingly, both places where interrogative words are used in English and other spoken languages. Thus, though the palm-up epistemic may be used for a wider set of meanings in sign than in gesture, these additional uses appear to be extensions of the interrogative meaning that is attested in gesture. In delving into the meaning of the palm-up epistemic in the next section, we focus on the six categories where there is clear attested overlap between gesture and sign.

On a cautionary note, there are limitations to the existing literature on the palm-up epistemic in sign, and these parallel the limitations of the gesture literature. For one, many of the observations collated above are drawn from brief mentions, and do not always include fine-grained descriptions of form. It is thus unclear whether the palm-up epistemic in sign resembles the prototypical form of the gesture discussed earlier—indeed, beyond the core palm-up aspect of the form, there appears to be considerable variation across languages (see **Figure 3**). Further, since interrogatives have become a topic of interest in sign language typology (Zeshan, 2004, 2006a), a number of sources comment on the palm-up in this context without venturing observations about wider usage. Finally, as in the gesture literature, there are only a handful of quantitative corpus treatments, making it difficult to assess, for instance, how commonly the palm-up is used to express the various meanings ascribed to it. Thus, as in gesture, further research is warranted.

THE PALM-UP PUZZLE

We now turn to the puzzle highlighted in our title. The broader puzzle concerns the meanings and origins of the entire palm-up form family. But a smaller and especially perplexing puzzle concerns the meanings and origins of the palm-up epistemic gesture in particular. This second, smaller puzzle has two parts. First, how are the six superficially distinct meanings for the palm-up epistemic gesture related, if indeed they are? Second, why is this form used for these meanings? In this section, we take these questions in turn.



How Are These Meanings Related?

Several of the meaning categories just discussed appear obviously connected, others less so. Why should we assume that these meanings are related in the first place? In making this assumption, we follow an inference commonly made in the study of linguistic polysemy: when one form covers the same meanings in different languages, there is most likely a conceptual link between those meanings, however distinct they may seem on the surface (e.g., Jurafsky, 1996; Evans and Wilkins, 2000). Indeed, so-called “accidental homophony” is usually considered an explanatory last resort. Given that the palm-up epistemic is associated with each of the six meaning categories in more than one community, we thus assume there are conceptual links between these meanings.

In trying to make sense of these links, we take inspiration from other accounts of cross-linguistic tendencies in meaning extension, such as Jurafsky’s (1996) account of the sprawling meanings of diminutives. Such accounts posit a core meaning and then show how other observed meanings can be understood as extensions from that core, or as extensions from extensions. Together these nodes and extensions comprise what might be

called a meaning network. We take the core meaning of the palm-up epistemic to be the expression of *absence*—of knowledge (“I don’t know”), ability (“I can’t”), or concern (“I don’t care”)—on the part of the speaker (for similar proposals, see Kendon, 1980; Zeshan, 2013). Importantly, the form is not widely used to convey the objective absence of some external entity, substance, or quality in the world (“There is none” or “There are none”) (but see footnote 2). Rather, it is used to convey the absence of some inner state or attitude. This meaning—which, for simplicity, we gloss as *absence of knowledge*—is among the most widely attested cross-linguistically, and each of the five other meaning categories ascribed to the palm-up can be considered extensions from this core. Here we discuss each of these extensions in turn, beginning with the more intuitive ones (e.g., how the *absence of knowledge* meaning motivates *uncertainty*-related meanings) and proceeding to the more surprising ones (e.g., how the *absence of knowledge* meaning motivates *exclamative* meanings). We summarize this meaning network in **Figure 4**, leaving aside for now meanings documented only in sign. Importantly, the fact that these extensions are attested across communities does not imply that the palm-up epistemic will exhibit *all* of these extensions in every community; it does imply, however, that communities will not skip over nodes in the network.

ABSENCE OF KNOWLEDGE > UNCERTAINTY

The expression of *absence of knowledge* can take different forms. Most basically, it can take the form of a confident statement that one lacks relevant knowledge, a simple assertion of “I don’t know.” However, language users are often unsure about whether they remember a fact correctly, fully grasp a concept, or completely agree with a statement, and, accordingly, they distance themselves from the truth of their statements. Such expressions of *uncertainty* can be conceptualized as a higher order absence of knowledge—that is, a lack of knowledge *about* one’s own knowledge or belief. Speakers have a range of resources for conveying uncertainty, including linguistic resources discussed under the banners of “modality” (Palmer, 1986) or “epistemic stance” (Du Bois, 2007), and a range of gestural resources beyond palm-ups (Roseano et al., 2014). In

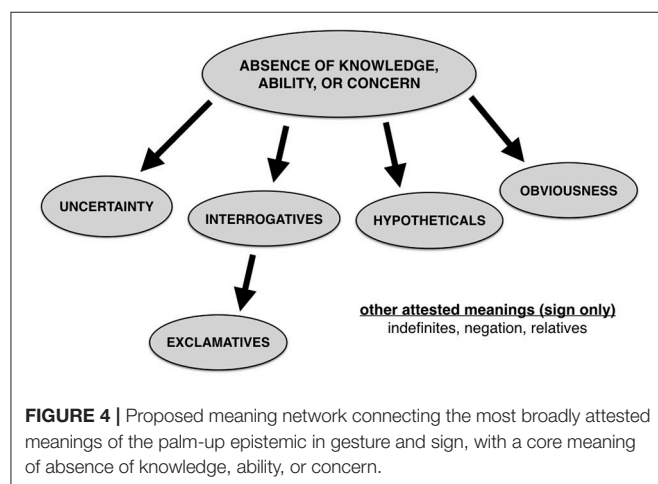
English, available linguistic resources include modal words like “maybe,” “perhaps,” or “could,” as well as so-called hedges, such as “I’m not sure,” “I guess,” “Well...” and so on. Linguistic hedges have been described as devices for distancing oneself from the truth (or falsity) of a proposition, giving language users the resources to express things that aren’t quite true, aren’t quite false, or aren’t quite true or false (Lakoff, 1973). Gestural hedges like the palm-up epistemic seem to perform the same function. The *uncertainty* category may partially account for why palm-ups are highly frequent in corpus studies of signed communication—the form, among its other functions, seems to be a favored pragmatic hedge in some sign languages (e.g., NZSL). Interestingly, instances of the palm-up sign used as a pragmatic hedge or hesitation marker tend to be classified as “gestural”—that is, part of signed communication but not part of the grammar or lexicon (e.g., McKee and Wallingford, 2011).

ABSENCE OF KNOWLEDGE > INTERROGATIVES

Another of the meaning categories most widely associated with the palm-up epistemic is interrogatives. While some authors describe the form as being associated with particular subtypes of questions (e.g., rhetorical; Kendon, 2004), most link it to the broader category of interrogatives. The link between *absence of knowledge* and *interrogative* meanings is perhaps intuitive. A question, after all, can be thought of as doing two things: first, implying that the questioner lacks relevant knowledge, and, second, putting it to the addressee to supply that knowledge (Wierzbicka, 1977; Kendon, 1980; Franklin et al., 2011). Thus, much as the gesture may be used when the speaker is expressing absence of knowledge, it may also be used when the speaker is both expressing absence of knowledge and asking the interlocutor to supply that knowledge. Interestingly, it is the interrogative uses of the palm-up epistemic that appear to be the best studied and documented in sign languages (see Zeshan, 2004; **Table 2**); but whether the interrogative use is indeed the one most often lexicalized across sign languages remains a question for future work.

ABSENCE OF KNOWLEDGE > HYPOTHETICALS

The *absence of knowledge* sense also extends to hypotheticals, such as those associated with the palm-up epistemic in Syuba (Gawne, 2018); statements in the subjunctive mood (“I wish...”), such as those associated with the palm-up particle in ASL (Conlin et al., 2003); and “whether” statements, such as those associated with the sign in Enga (Kendon, 1980). For simplicity—and lack of an appropriate general term—we refer to these uses together as “hypotheticals.” The link to *absence of knowledge* is, again, relatively intuitive: when speakers describe a state of affairs that has not happened and may or may not happen, they implicitly convey an absence of knowledge about that state of affairs. Interestingly, these and other uses of the palm-up epistemic appear to fall under the umbrella of *irrealis* (e.g., Elliott, 2000). This is a broad category, covering statements of all kinds about events or facts that are not “real” in the sense that they have not yet happened; in some accounts, it



includes hypotheticals, interrogatives, imperatives, and more. But, importantly, the palm-up epistemic does not appear to be associated with the entire *irrealis* category—for example, there is no evidence for an association between palm-ups and imperatives.

ABSENCE OF KNOWLEDGE > OBVIOUSNESS

A meaning category that is less intuitively related to the *absence of knowledge* is *obviousness*. Indeed, this extension is, at first blush, puzzling: Why would the very same gestural form sometimes be used to convey a lack of certainty and, other times, to convey a conviction that something is so certain as to be obvious? One straightforward account of this link is that expressing that something is obvious amounts to expressing that “I don’t know what else I could say about it” (Jehoul et al., 2017, p. 7). The use of the palm-up epistemic to express obviousness resembles a similarly counter-intuitive extension of gestural meaning observed in the case of headshakes: speakers commonly shake their heads while making extreme *positive* evaluations (e.g., “It was marvelous”; Kendon, 2002, p. 172–3). A possible explanation is that, in such cases, the speaker is rejecting an implicit assumption that something is ordinary or unremarkable. In a similar way, when using palm-ups to express obviousness, speakers may be reacting to an implicit assumption that more could or should be said—they are asserting that, in fact, they do not know more, do not care more, or are not able to say more. Another account of the link between absence of knowledge and obviousness would consider it a less-direct extension, mediated by interrogative uses of the palm-up epistemic. On this account, the statement that something is obvious can be seen as entailing an implicit question, such as “How could it be otherwise?” “How could you not know this?” or “What else could one say?” More data are needed to adjudicate between these possible paths of extension; for now, we default to the more parsimonious assumption that *obviousness* extends directly from *absence of knowledge*. Regardless of the extension path, this use of the gesture is distinct from the others discussed so far in that it expresses something about the speaker’s *affective* state. Here the palm-up serves what is sometimes described as an *expressive* function (e.g., Cruse, 1986; Potts, 2007) in that it changes the affective coloring of the utterance but not the information it conveys—that is, its assertive content.

INTERROGATIVES > EXCLAMATIVES

Another meaning category less obviously connected to the others is *exclamatives*. *Exclamatives* are statements exhibiting a high degree of affect, whether positive or negative (in this category we include uses of palm-ups as part of “emphatic statements”; Rector, 1986; Conlin et al., 2003). As with the category of obviousness, there is something initially puzzling here. Why would the same gesture be sometimes used to convey a lack of certainty or concern and, other times, to convey extreme certainty or concern? And, again, as with the category of obviousness, exclamative uses of the palm-up epistemic are fundamentally *expressive*. We interpret the association of palm-ups with *exclamatives* as an extension of their association with *interrogatives*. This extension path parallels the cross-linguistically robust phenomenon in spoken languages

whereby interrogative words are used to form exclamatives (e.g., Bolinger, 1972; Wierzbicka, 1977; Espinal, 1995). Examples in English include expressions such as “How rude!” and “What a jerk!” Further, though cross-linguistically less common (Rett, 2008), exclamatives may also be derived from polar interrogatives (e.g., “Boy, did she run!”). The precise semantic-pragmatic motivation for this repurposing of interrogative structures in exclamatives remains a matter of theoretical discussion (e.g., Rett, 2008). These clear links to interrogatives notwithstanding, it should also be noted that exclamatives can be marked in a number of ways—that is, utterances with exclamative force are not uniformly couched in a particular structure. In a similar way, the palm-up epistemic gesture appears to be associated with exclamations *generally* (e.g., “That’s great!”), whether or not they involve interrogative words or other interrogative structures.

Evaluating the Proposed Extension Paths

The meaning network just proposed crystallizes a hypothesis, one that remains to be tested and refined. Doing so will require more data—in particular, more detailed, systematic, quantitative analyses from across languages, both spoken and signed. Here we highlight several kinds of data that would be especially useful in assembling a clearer picture.

A first type of data that would be useful are observations over the lifespan, that is, developmental data. Knowing how children use the palm-up epistemic gesture initially, for instance, may shed light on its core meaning. Though we have proposed that the core of the gesture is *absence of knowledge* (see also Kendon, 1980; Zeshan, 2013), there are other possibilities. For instance, the gesture could have roots in the expression of external, objective absence, rather than absence of knowledge, ability, or concern. The emblematic “all gone” gesture—used to remark on objective absence—is well attested in children’s early communication and in child-directed speech (see footnote 2). Indeed, some observations suggest that this gesture may emerge before epistemic uses of the palm-up form (e.g., Beaupoil-Hourdel and Debras, 2017). Whether such observations contradict our proposal, however, is unclear. The distribution of the “all gone” gesture is currently unknown. Such a convention may occasionally arise because *objective absence* is a more accessible meaning for young children than *absence of knowledge*. That is, the “all gone” gesture could be a kind of a gestural “back formation” that becomes conventionalized in some communities. More cross-linguistic developmental data will be needed to explore this possibility.

Developmental data would also shed light on particular paths of meaning extension. We would not necessarily expect children to use the palm-up for all of the attested meanings in the network—much as we would not expect all communities to use the palm-up for all meanings in the network—but, again, we would expect children not to skip over nodes in the network. Thus, if our proposed path from *absence of knowledge* to *interrogatives* to *exclamatives* is correct, children may not use the palm-up epistemic exclamatively until they have already begun to use it interrogatively; in turn, they may not use it interrogatively until they have already begun to use it to express a lack of knowledge. In practice, such semantic extensions may be hard

to detect because several meanings may emerge within a narrow time frame. Developmental analyses of this sort have recently begun to shed light on how other bodily forms of communication take on new meanings (for the case of negation, see André, 2014; Beaupoil-Hourdel et al., 2016). And one recent study throws some valuable initial light on developmental changes in how palm-up forms are used. Graziano (2014) examined the emergence of “palm lateral” (in our terms, the palm-up epistemic) and “palm presentation” (in our terms, the palm-up presentational) gestures in Italian children between the ages of 4 and 10. She found that palm-up epistemic gestures were present in the youngest children, but that palm-up presentational gestures did not emerge until later ages. Moreover, she noted that children first used the palm-up epistemic along with “crystallized expressions” (p. 311), such as “I don’t know,” and only later used them more flexibly as adults do, e.g., to express obviousness. This finding is consistent with our suggestion that obviousness is an extension of absence of knowledge, and thus should emerge later. Further studies of the developmental changes in use of the palm-up form family would be valuable, including studies in different speaking communities and with even younger children.

Another important source of data would be additional studies with adult speakers and signers, both corpus-based and experimental. A corpus-based analysis using the categories of meanings described above, for instance, could shed light on which meanings are most common within and across languages. At present, our understanding of the relative prominence of these different meanings is sketchy at best, based largely on the number of communities in which they have been reported. Corpus studies may also reveal additional recurring meanings of the palm-up epistemic beyond the six we have focused on. There have already been several insightful corpus-based treatments of the palm-up in sign, but especially valuable would be further studies that compare use of the form in different sign languages using the same analytic criteria and theoretical framework. Such an approach would be critical in distinguishing cross-linguistic patterns from language-specific particulars.

Experimental studies in both speakers and signers would provide complementary insights. Elicitation tasks would be helpful in discerning the strength of association between particular meaning categories and the palm-up epistemic. In gesturers, a well-devised elicitation task might tell us whether, for instance, speakers associate the gesture more strongly with expressions of *absence of knowledge* than with expressions of *obviousness*, as might be predicted from the fact that *absence of knowledge* is the proposed core. In signers, similar tasks could shed light on which uses of the palm-up epistemic are strongly tied to certain contexts—and thus, by hypothesis, are more grammaticalized—and which are less strongly tied—and thus are more gestural, or affective. Judgment tasks with both groups could also be illuminating. Do listeners find palm-up epistemics in certain discourse contexts—or co-produced with certain words (e.g., interrogative words)—to be more natural than others? Such studies could shed crucial light on the shadings of meaning that palm-ups add when conjoined with certain kinds of discourse content or when produced in certain conversational positions.

Why This Form for These Meanings?

The second part of the palm-up puzzle is why these meanings are associated with this form in particular. We assume there is indeed a motivation behind the pairing of these meanings with this form simply because of the recurrence of the pairing across communities. This inductive inference is parallel to one commonly made in studies of figurative language and grammaticalization: if the same target concept is expressed using the same source concept in more than one speech community, there is likely a motivated relationship between the two concepts (e.g., Brown and Witkowski, 1981; Sweetser, 1990; Heine, 1997). But it is also possible, of course, that there really is no motivation to explain. On this skeptical account, the palm-up epistemic could be merely a “catchy convention” that has spread far and wide, emanating perhaps from some centuries-old source in European culture. We think this scenario is unlikely. The wide distribution of the gesture—a distribution matched only by a few other bodily communicative forms, such as headshakes, index-finger pointing, and certain facial expressions—suggests independent development in different communities. And independent development, in turn, suggests an underlying motivation. There can be little doubt that there are conventional *aspects* to the palm-up. That is, part of why people use it in the ways that they do (e.g., with the distinctive handshape used in Syuba) is that others in their community use it in these ways. Crucially, however, just because a communicative form has conventional aspects does not mean it is unmotivated (e.g., Jakobson, 1972; Enfield, 2009). Here we consider two explanations for the underlying motivation between the palm-up form and the meaning it is so widely associated with—*absence of knowledge*, which we take to be its core meaning. As will become clear, questions about the origins of palm-up gestures are impossible to separate from a sticking point with which we started: whether there is one interconnected family of palm-up gestures or separate gestures with similar forms.

The Metaphorical Account

A first of account of the origins of the palm-up epistemic gesture is that it is a kind of metaphorical action, rooted in practical actions of giving and receiving physical objects (e.g., Müller, 2004; see also McNeill, 1992; Engberg-Pedersen, 2002; Parrill, 2008; Streeck, 2009). On this account, the gesture expresses the “conduit metaphor” (Reddy, 1979; McNeill, 1992), in which discourse is understood as object exchange: ideas are presented and requested much as real objects are in routine activities. In line with this metaphor, palm-ups generally can be seen as representing that the hands are metaphorically full, and a discourse object is being offered to the listener, or that the hands are metaphorically empty, and a discourse object is being requested. Intuitively, if palm-ups sometimes represent empty hands, we have a plausible motivation for what we take to be the core meaning of the palm-up epistemic: that the speaker *lacks* some knowledge, internal state, or attitude (e.g., Chu et al., 2013). In a similar way, when using the palm-up in the course of asking a question, the speaker may be showing that the hand is “empty” of knowledge, or, perhaps, inviting the listener to put some knowledge into that empty hand (e.g., Müller, 2004).

Kendon (2004) also describes the gesture as rooted in practical action but keys in on a different aspect of the gesture's form: he sees the lateral movement as indicating that "whatever has been presented is being withdrawn from" (p. 265). There are certainly appealing features of this type of metaphorical account. The general idea that many discourse-related gestures are rooted in practical actions has explanatory power and intuitive plausibility. In the case of the headshake used for negation, for instance, researchers have made a plausible case that its motivations lie in the practical action of averting the face from a food source (Darwin, 1998/1872; Beaupoil-Hourdel et al., 2016), and many recurrent gestures seem to be related to action schemas (e.g., Müller, 2017).

However, there are also limitations to such a metaphorical account as it applies to palm-ups. Most importantly, the account is mum about the widely observed relation between palm-up gestures and shrugs. As has been widely observed, palm-ups and shrugs are frequently co-produced and overlap in meaning (as noted in Kendon, 1980, 2004; Franklin et al., 2011; Chu et al., 2013). And, yet, people do not shrug their shoulders as part of the routine exchange of objects. The shrug seems to demand an explanation outside of the realm of practical action, and so, too, may the palm-up epistemic. The second explanation for the gesture's motivation centers on its relation to the shrug.

The Reduced Shrug Account

A different account sees the palm-up epistemic as derived from the shrug. The shrug—as described originally by Darwin (1998/1872) and by several observers since (Givens, 1977, 2016; Streeck, 2009; Debras, 2017; Jehoul et al., 2017)—is a multifaceted display that very often involves rotating the forearms so that the palms turn upward. It has been attested in a range of cultures, in both speakers (e.g., Creider, 1977; Feldman, 1986; Agwuele, 2014) and signers (e.g., Zeshan, 2006b; McKee and Wallingford, 2011; Schuit, 2014), and is sometimes considered a "candidate gesture universal" (Streeck, 2009, p. 189). Interestingly, the meanings of the shrug are less controversial than the meanings of palm-ups. Observers broadly agree that the shrug is used primarily to indicate absence of knowledge, ability, or concern—the same meaning we have described as the core meaning of the palm-up epistemic gesture—and that it can also be used to indicate both uncertainty and obviousness (Debras, 2017; Jehoul et al., 2017), two of the other meaning categories commonly associated with the palm-up epistemic⁶. But what motivates the connection between the shrug and these meanings? As noted already, the shrug is not a component of practical actions involving object exchange. Darwin (1998/1872) explained its origins in a different way, by invoking his "principle of antithesis." According to this principle, a certain meaning will sometimes be expressed with a certain bodily form, not because that bodily form itself naturally expresses the meaning, but because it *contrasts* with another bodily form that naturally expresses the *contrasting* meaning. In

other words, a form of expression may be motivated in that its "antithesis" is motivated. The widespread use of the head nod for affirmation may be described in just this way: it is not naturally connected to affirmation, but in its vertical movement pattern it contrasts with the lateral pattern of the head shake, which many have argued is naturally connected to negation because it emerges out of the act of refusing food. For Darwin, to understand why the shrug means what it does we must first understand its bodily opposite: an aggressive, fighting stance, which involves making fists, squaring the shoulders, and making the arms rigid. By bodily contrast with this assertive posture, the shrug embodies a non-assertive, non-aggressive attitude. A related proposal is that the shrug is rooted in a primordial crouching posture (Givens, 1986). Both explanations highlight the non-assertiveness of the shrug, and might thus plausibly account for why the gesture would be used widely to express *absence of knowledge, ability, or concern* and, by extension, the other meanings reviewed earlier.

As others have noted, the palm-up epistemic when produced on its own can be described as a reduced, or manual-only version of the shrug. Darwin (1998/1872) himself noted that shoulder movement is an optional component; he describes, for instance, a shrug that takes the form of a "mere turning slightly outwards of the open hands" (p. 266). One group of researchers refers to palm-ups, in fact, as "hand shrugs" (Johnson et al., 1975). Another group, in investigating the prevalence of different bodily components of the shrug display, reports that the palm-up form is, in fact, a more frequent component of the shrug than shoulder action (Jehoul et al., 2017, p. 3). More generally, many researchers since Darwin have noted connections between the shrug and the palm-up (Givens, 2016; Debras, 2017), and some have described them as functionally interchangeable (Chu et al., 2013). Regardless of whether one endorses Darwin's account of the antithetical origins of the shrug in its particulars, a strength of the reduced shrug account is that it takes seriously the clear affinity between palm-up epistemics and shrugs, rather than ignores it.

But the reduced shrug account is not without limitations. Notably, the meanings of the shrug and of the palm-up epistemic, while clearly overlapping, do not appear to be completely co-extensive. Absence of knowledge, uncertainty, and obviousness have all been attributed to both gestures. But, notably, shrugs do not appear to be widely used for interrogative functions (though see Franklin et al., 2011). How might we account for this partial dissociation? A possible explanation concerns how readily different bodily actions can be produced to overlap with speech. Shoulder shrugs with palm-ups, shoulder shrugs without palm-ups, and palm-ups without shoulder shrugs can all be used to good effect on their own—that is, without co-occurring speech, or as prefaces or end-caps on utterances. However, the palm-up and shrug are not equally suited to spanning over long stretches of talk. In order to be salient, the shoulder-hunching component of the shrug requires a relatively quick up-and-down movement of the shoulders. Shoulder shrugs are thus not easily held in a way that spans across speech. Palm-ups, by contrast, do not have this limitation, as they remain salient when held. Intriguingly, one recent study noted a marked difference in how long different components of the shrug (e.g., shoulder action, palm-up form,

⁶Other uses of the shrug have also been described. Calbris (2011) notes that in French it can be used to signify: exclamation; powerlessness; or getting rid of something insignificant. She sees the form of the shrug as motivated by the idea of removing an "annoying object" through the shoulder movement.

and head tilt) were held (Jehoul et al., 2017, p. 8). Speculatively, if the palm-up form is a component of the full-blown shrug display better suited for spanning over speech, it may be more strongly associated with functions that take scope over an utterance (e.g., interrogatives).

To be sure, the ultimate origins of this form-meaning pairing are hard to pin down decisively. Part of the difficulty in adjudicating between the “metaphorical” and “reduced shrug” accounts just sketched is that they tend to have different explanatory targets. Many proponents of the metaphorical account do not observe a distinction, as we and others do (Kendon, 2004; Chu et al., 2013; Graziano, 2014), between what we have called palm-up presentational gestures and palm-up epistemic gestures. Rather, they have in mind a broader family of palm-up forms with a broader family of meanings, built around a “presentational” core (Engberg-Pedersen, 2002; Müller, 2004; Parrill, 2008). A compelling possibility, in our view, is that the metaphorical account best explains the palm-up presentational, whereas the reduced shrug account best explains the palm-up epistemic. On this account, the palm-up presentational and palm-up epistemic gestures may be best thought of as “false friends”—that is, communicative forms that look deceptively alike, but actually have quite different meanings and origins. But, certainly, to begin to adjudicate between possible origin stories, we first need a better handle on the relationship between these two proposed gesture variants. Other observations could also tip the balance in favor of one or the other of the origin stories outlined above. If the reduced shrug account is correct, we would not expect to find the palm-up epistemic gesture in broad use except in those cultures where the shrug is also in broad use. We might also expect to find that, developmentally, the shrug precedes, or at least co-occurs with, the palm-up epistemic. The metaphorical account does not predict either of these patterns.

CONCLUSION

One of the most common gestural forms that speakers produce in everyday communication involves rotating the forearms so that the palms face upward. Palm-up forms are captured in the paintings of Renaissance masters and in the GIFs and emoji of contemporary social media; they are produced by gesturers and signers the world over. In their pervasiveness, cross-linguistic spread, and frequent incorporation into sign language grammar, palm-up forms may be surpassed only by pointing and head gestures. And yet palm-ups remain puzzling. They vary considerably from one use to the next, even in sign languages; they go by different labels; they resist current gesture classification schemes and elude existing linguistic categories. In

fact, it is not even clear what the palm-up form family consists of—one sprawling family of interconnected meanings, a family with salient divisions, or perhaps a pair of “false friends.” A number of meanings have been attributed to palm-ups, not always obviously connected to each other, and sometimes even contradictory. And fundamentally different accounts have tried to explain the fact that this particular form—however we label, classify, or circumscribe it—is used for similar meanings in culture after culture.

Here, we have tried to find some clarity amid these complexities. Following others, we proposed a distinction between the palm-up presentational and the palm-up epistemic, and focused our attention on the latter. We showed that, in the existing literature, at least six meaning categories have been recurrently associated with this variant of the palm-up in both gesture and sign. Examining what we have described as the first part of the palm-up puzzle—how these meanings are related—we showed that this set can be understood as extensions from a kernel meaning of *absence of knowledge*. Examining the second part—why this form is associated with this set of meanings—we sketched two accounts, a “metaphorical account” and a “reduced shrug account.” This does not mean, of course, that we can now pronounce the palm-up puzzle solved. But the first step in solving any puzzle is to figure out what the pieces are—and we hope to have made progress toward this more modest goal. We have also made concrete suggestions for where research on palm-ups could go next. Part of our interest in this puzzle has little to do with the palm-up form family *per se*. Rather, it has to do with meaning generally: with how bodily forms come to express abstract meanings; how meanings extend to new meanings; and how bodily forms combine with language—as in the case of co-speech or co-sign gestures—or even become grammaticalized—as in the case of sign. Future efforts to illuminate palm-ups will throw much-needed light on this broader puzzle too.

AUTHOR CONTRIBUTIONS

All authors discussed the literature reviewed and designed the content and structure of the article. KC drafted the manuscript, with critical revisions provided by NA and SG-M. All authors approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2018.00023/full#supplementary-material>

REFERENCES

- Aboh, E., Pfau, R., and Zeshan, U. (2005). “When a wh-word is not a wh-word: the case of Indian Sign Language,” in *The Yearbook of South Asian Languages and Linguistics*, ed R. Singh (Berlin: Walter de Gruyter), 11–43.
- Agwuele, A. (2014). A repertoire of Yoruba hand and face gestures. *Gesture* 14, 70–96. doi: 10.1075/gest.14.1.04agw
- Andrén, M. (2014). Multimodal constructions in children: is the headshake part of language? *Gesture* 2, 141–170. doi: 10.1075/gest.14.2.02
- Aronoff, M., Meir, I., Padden, C., and Sandler, W. (2005). “Morphological universals and the sign language type,” in *Yearbook of Morphology 2004*, eds G. Booij and J. van Marle (Boston, MA: Kluwer Academic Publishers), 9–39.
- Barakat, R. A. (1973). Arabic gestures. *J. Popul. Cult.* 6, 749–792. doi: 10.1111/j.0022-3840.1973.00749.x

- Bauer, A. (2015). *The Use of Signing Space in a Shared Sign Language of Australia*. Berlin: Walter de Gruyter.
- Bavelas, J. B., Chovil, N., Lawrie, D. A., and Wade, A. (1992). Interactive gestures. *Discourse Process* 15, 469–489. doi: 10.1080/01638539209544823
- Beaupoil-Hourdel, P., and Debras, C. (2017). Developing communicative postures: the emergence of shrugging in child communication. *Lang. Interact. Acquis.* 8, 89–116. doi: 10.1075/lia.8.1.05bea
- Beaupoil-Hourdel, P., Morgenstern, A., and Boutet, D. (2016). “A child’s multimodal negations from 1 to 4: the interplay of modalities,” in *Negation and Polarity: Experimental Perspectives*, eds P. Larrivée and C. Lee (Cham: Springer), 95–123.
- Bolinger, D. (1972). *Degree Words*. Paris: Mouton.
- Börstell, C., Hörberg, T., and Östling, R. (2016). Distribution and duration of signs and parts of speech in Swedish Sign Language. *Sign Lang. Linguist.* 19, 143–196. doi: 10.1075/sll.19.2.01bor
- Brookes, H. (2004). A repertoire of South African quotable gestures. *J. Linguist. Anthropol.* 14, 186–224. doi: 10.1525/jlin.2004.14.2.186
- Brown, C. H., and Witkowski, S. R. (1981). Figurative language in a universalist perspective. *Am. Ethnol.* 8, 596–615. doi: 10.1525/ae.1981.8.3.02a00110
- Calbris, G. (1990). *The Semiotics of French Gestures*. Bloomington, IN: Indiana University Press.
- Calbris, G. (2011). *Elements of Meaning in Gesture*. Philadelphia, PA: John Benjamins.
- Chu, M., Meyer, A., Foulkes, L., and Kita, S. (2013). Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy. *J. Exp. Psychol. Gen.* 143, 694–709. doi: 10.1037/a0033861
- Conlin, F., Hagstrom, P., and Neidle, C. (2003). A particle of indefiniteness in ASL. *Linguist. Discov.* 2, 1–21. doi: 10.1349/PS1.1537-0852.A.142
- Cooperrider, K., and Núñez, R. (2012). Nose-pointing: notes on a facial gesture of Papua New Guinea. *Gesture* 12, 103–129. doi: 10.1075/gest.12.2.01coo
- Creider, C. A. (1977). Towards a description of East African gestures. *Sign Lang. Stud.* 14, 1–20. doi: 10.1353/sls.1977.0014
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
- Darwin, C. (1998/1872). *The Expression of the Emotions in Man and Animals*. New York, NY: Oxford University Press.
- Debras, C. (2017). The shrug: forms and meanings of a compound enactment. *Gesture* 16, 1–34. doi: 10.1075/gest.16.1.01deb
- Du Bois, J. W. (2007). “The stance triangle,” in *Stancetaking in Discourse*, ed R. Englebretson (Philadelphia, PA: John Benjamins), 139–182.
- Ekman, P., and Friesen, W. V. (1972). Hand movements. *J. Commun.* 22, 353–374. doi: 10.1111/j.1460-2466.1972.tb00163.x
- Elliott, J. R. (2000). Realis and irrealis: forms and concepts of the grammaticalisation of reality. *Linguist. Typol.* 4, 55–90. doi: 10.1515/lity.2000.4.1.55
- Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge: Cambridge University Press.
- Engberg-Pedersen, E. (2002). “Gesture in signing: the presentation gesture in Danish Sign Language,” in *Progress in Sign Language Research: in Honor of Siegmund Prilwitz*, eds R. Schulmeister and H. Reinitzer (Hamburg: Signum), 143–162.
- Espinal, M. T. (1995). “Non-negative negation and wh-exclamatives,” in *Negation and Polarity*, eds D. Forget, P. Hirschbühler, F. Martineau, and M.-L. Rivero (Amsterdam: John Benjamins), 75–93.
- Evans, N., and Wilkins, D. (2000). In the mind’s ear: the semantic extensions of perception verbs in Australian languages. *Language* 76, 546–592. doi: 10.2307/417135
- Feldman, H. (1986). *A Grammar of Awtuw*. Canberra, ACT: Australian National University.
- Fenlon, J., Schembri, A., Rentelis, R., Vinson, D., and Cormier, K. (2014). Using conversational data to determine lexical frequency in British Sign Language: the influence of text type. *Lingua* 143, 187–202. doi: 10.1016/j.lingua.2014.02.003
- Ferré, G. (2011). Functions of three open-palm hand gestures. *Multimodal Commun.* 1, 5–20. Available online at: <https://hal.archives-ouvertes.fr/hal-00666025>
- Franklin, A., Giannakidou, A., and Goldin-Meadow, S. (2011). Negation, questions, and structure building in a homesign system. *Cognition* 118, 398–416. doi: 10.1016/j.cognition.2010.08.017
- Gawne, L. (2018). Contexts of use of a rotated palms gesture among Syuba (Kagate) speakers in Nepal. *Gesture* 17, 37–64. Available online at: <https://benjamins.com/catalog/gest.00010.gaw>
- Givens, D. B. (1977). Shoulder-shrugging: a densely communicative expressive behavior. *Semiotica* 19, 13–28. doi: 10.1515/semi.1977.19.1-2.13
- Givens, D. B. (1986). The big and the small: toward a paleontology of gesture. *Sign Lang. Stud.* 51, 145–170. doi: 10.1353/sls.1986.0024
- Givens, D. B. (2016). Reading palm-up signs: neurosemiotic overview of a common hand gesture. *Semiotica* 2016, 235–250. doi: 10.1515/sem-2016-0053
- Goldin-Meadow, S. (2003). *The Resilience of Language: What Gesture Creation in Deaf Children Can Tell Us About How All Children Learn Language*. New York, NY: Psychology Press.
- Goldin-Meadow, S., and Brentari, D. (2017). Gesture, sign and language: the coming of age of sign language and gesture studies. *Behav. Brain Sci.* 40, 1–45. doi: 10.1017/S0140525X15001247
- Goldin-Meadow, S., and Mylander, C. (1984). Gestural communication in deaf children: the effects and noneffects of parental put on early language development. *Monogr. Soc. Res. Child Dev.* 49, 1–151. doi: 10.2307/1165838
- Graziano, M. (2014). “The development of two pragmatic gestures of the so-called Open Hand Supine family in Italian children,” in *From Gesture in Conversation to Visible Action as Utterance*, eds M. Seyfeddinipur and M. Gullberg (Philadelphia, PA: John Benjamins), 311–330.
- Grice, H. P. (1975). “Logic and conversation,” in *Syntax and Semantics, Vol. 3, Speech Acts*, eds P. Cole and J. L. Morgan (New York, NY: Academic Press), 41–58.
- Haspelmath, M. (1997). *Indefinite Pronouns*. Oxford: Clarendon Press.
- Haviland, J. B. (2015). Hey! *Top. Cogn. Sci.* 7, 124–149. doi: 10.1111/tops.12126
- Heine, B. (1997). *Cognitive Foundations of Grammar*. Oxford: Oxford University Press.
- Isaacson, W. (2017). *Leonardo Da Vinci*. New York, NY: Simon & Schuster.
- Iverson, J. M., Capirci, O., Volterra, V., and Goldin-Meadow, S. (2008). Learning to talk in a gesture-rich world: early communication in Italian vs. American children. *First Lang.* 28, 164–181. doi: 10.1177/0142723707087736
- Jakobson, R. (1972). Motor signs for “yes” and “no.” *Lang. Soc.* 1, 91–96.
- Jehoul, A., Brône, G., and Feyaerts, K. (2017). The shrug as marker of obviousness. *Linguist. Vanguard* 3:20160082. doi: 10.1515/lingvan-2016-0082
- Johnson, H. G., Ekman, P., and Friesen, W. V. (1975). Communicative body movements: American emblems. *Semiotica* 15, 335–353. doi: 10.1515/semi.1975.15.4.335
- Johnston, T. (2012). Lexical frequency in sign languages. *J. Deaf Stud. Deaf Educ.* 17, 163–193. doi: 10.1093/deafed/enr036
- Jurafsky, D. (1996). Universal tendencies in the semantics of the diminutive. *Language* 72, 533–578. doi: 10.2307/416278
- Kendon, A. (1980). A description of a deaf-mute sign language from the Enga Province of Papua New Guinea with some comparative discussion. Part III: Aspects of utterance construction. *Semiotica* 32, 245–313.
- Kendon, A. (2002). Some uses of the head shake. *Gesture* 2, 147–182. doi: 10.1075/gest.2.2.03ken
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Lakoff, G. (1973). Hedges: a study in meaning criteria and the logic of fuzzy concepts. *J. Philos. Logic* 2, 458–508.
- McGowan, T. (2010). Abstract deictic gestures-in-interaction: a barometer of intersubjective knowledge development in small-group discussion. *Work. Pap. Edu. Linguist.* 25, 55–79.
- McKee, R. L., and Wallingford, S. (2011). “So, well, whatever”: discourse functions of palm-up in New Zealand Sign Language. *Sign Lang. Linguist.* 14, 213–247. doi: 10.1075/sll.14.2.01mck
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychol. Rev.* 92, 350–371.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: Chicago University Press.
- Meir, I. (2004). Question and negation in Israeli Sign Language. *Sign Lang. Linguist.* 7, 97–124. doi: 10.1075/sll.7.2.03mei
- Monahan, B. (1983). *A Dictionary of Russian Gesture*. Ann Arbor, MI: Hermitage.
- Müller, C. (2004). “Forms and uses of the Palm Up Open Hand: a case of a gesture family?” in *The Semantics and Pragmatics of Everyday Gestures*, eds C. Müller and R. Posner (Berlin: Weidler), 233–256.

- Müller, C. (2017). How recurrent gestures mean and embodied motivation. *Gesture* 16, 276–303. doi: 10.1075/gest.16.2.05mul
- Nyst, V. (2012). “Shared sign languages,” in *Sign Language*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 552–574.
- Palmer, F. R. (1986). *Mood and Modality*. Cambridge: Cambridge University Press.
- Parrill, F. (2008). “Form, meaning, and convention: a comparison of a metaphoric gesture with an emblem,” in *Metaphor and Gesture*, eds A. Cienki and C. Müller (Amsterdam: John Benjamins), 195–217.
- Payrató, L. (1993). A pragmatic view on autonomous gestures: a first repertoire of Catalan emblems. *J. Pragmat.* 20, 193–216. doi: 10.1016/0378-2166(93)90046-R
- Pfau, R., and Steinbach, M. (2006). Modality-independent and modality-specific aspects of grammaticalization in sign languages. *Linguist. Potsdam* 24, 135–182.
- Potts, C. (2007). The expressive dimension. *Theor. Linguist.* 33, 165–198. doi: 10.1515/TL.2007.011
- Rector, M. (1986). “Emblems in Brazilian culture,” in *Iconicity: Essays on the Nature of Culture*, eds P. Bouissac, M. Herzfeld, and R. Posner (Tübingen: Stuaffenburg), 447–462.
- Reddy, M. J. (1979). “The conduit metaphor: a case of frame conflict in our language about language” in *Metaphor and Thought*, ed A. Ortony (Cambridge: Cambridge University Press), 164–201.
- Rett, J. (2008). A degree account of exclamatives. *Seman. Linguist. Theor.* 18, 601–618. doi: 10.3765/salt.v18i0.2470
- Roseano, P., González, M., Borrás-Comes, J., and Prieto, P. (2014). Communicating epistemic stance: how speech and gesture patterns reflect epistemicity and evidentiality. *Discourse Process.* 53, 135–174. doi: 10.1080/0163853X.2014.969137
- Saitz, R. L., and Cervenká, E. J. (1972). *Handbook of Gestures: Columbia and the United States*. The Hague: Mouton.
- Schuit, J. M. (2014). *Signs of the arctic: Typological aspects of Inuit Sign Language*. Doctoral dissertation, Print Service Ede.
- Sherzer, J. (1973). Verbal and nonverbal deixis: the pointed lip gesture among the San Blas Cuna. *Lang. Soc.* 2, 117–131. doi: 10.1017/S0047404500000087
- Shuman, M. K., and Cherry-Shuman, M. M. (1981). A brief annotated sign list of Yucatec Maya sign language. *Lang. Sci.* 3, 124–185. doi: 10.1016/S0388-0001(81)80017-4
- Spaepen, E., Coppola, M., Flaherty, M., Spelke, E., and Goldin-Meadow, S. (2013). Generating a lexicon without a language model: do words for number count? *J. Mem. Lang.* 69, 496–505. doi: 10.1016/j.jml.2013.05.004
- Sparhawk, C. M. (1978). Contrastive-identificational features of Persian gestures. *Semiotica* 24, 49–86. doi: 10.1515/semi.1978.24.1-2.49
- Streeck, J. (2003). A moment and its gestures: a discussion of Leo Steinberg’s analysis of Leonardo’s Last Supper. *Gesture* 3, 213–236. doi: 10.1075/gest.3.2.06str
- Streeck, J. (2009). *Gesturecraft: The Manu-facture of Meaning*. Amsterdam: John Benjamins.
- Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantics*. Cambridge: Cambridge University Press.
- Ullman, R. (1978). “Some general characteristics of interrogative systems,” in *Universals of Human Language, Vol. 4, Syntax*, ed J. H. Greenberg (Stanford, CA: Stanford University Press), 211–248.
- Van Loon, E. (2012). *What’s in the Palm of Your Hand? Discourse Functions of the Palm-Up in Sign Language of the Netherlands*. Master’s thesis. University of Amsterdam.
- Van Loon, E., Pfau, R., and Steinbach, M. (2014). “The grammaticalization of gestures in sign languages,” in *Body-Language-Communication, Vol. 2*, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and J. Bressemer (New York, NY: Mouton de Gruyter) 720–730.
- Volk, E. (2017). “Palm-up: It’s not all about give and take,” in *Poster Session Presented at Meeting of Formal and Experimental Advances in Sign Language Theory* (Reykjavik).
- Washbaugh, W., Woodward, J. C., and DeSantis, S. (1978). Providence island sign language: a context-dependent language. *Anthropol. Linguist.* 20, 95–109.
- Wierzbicka, A. (1977). The ignorative: the semantics of speech acts. *Int. Rev. Slavic Linguist.* 2, 251–312.
- Wilbur, R. B. (2010). “The semantics-phonology interface,” in *Sign Languages*, ed D. Brentari (Cambridge: Cambridge University Press), 357–382.
- Wilcox, S. (2004). Gesture and language: cross-linguistic and historical data from signed languages. *Gesture* 4, 43–73. doi: 10.1075/gest.4.1.04wil
- Zeshan, U. (2003). Indo-Pakistani sign language grammar: a typological outline. *Sign Lang. Stud.* 3, 157–212. doi: 10.1353/sls.2003.0005
- Zeshan, U. (2004). Interrogative constructions in signed languages: crosslinguistic perspectives. *Language* 80, 7–39. doi: 10.1353/lan.2004.0050
- Zeshan, U. (ed.). (2006a). *Interrogative and Negative Constructions in Sign Languages*. Nijmegen: Ishara Press.
- Zeshan, U. (2006b). “Negative and interrogative structures in Turkish Sign Language (TID),” in *Interrogative and Negative Constructions in Sign Languages*, ed U. Zeshan (Nijmegen: Ishara Press), 128–164.
- Zeshan, U. (2013). “Question particles in sign languages,” in *The World Atlas of Language Structures Online*, eds M. S. Dryer and M. Haspelmath (Leipzig: Max Planck Institute for Evolutionary Anthropology). Available online at: <http://wals.info/chapter/140> (Accessed December 07, 2017).

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MG and handling Editor declared their shared affiliation.

Copyright © 2018 Cooperrider, Abner and Goldin-Meadow. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Iconicity in Signed and Spoken Vocabulary: A Comparison Between American Sign Language, British Sign Language, English, and Spanish

Marcus Perlman^{1*}, Hannah Little², Bill Thompson³ and Robin L. Thompson⁴

¹ Department of English Language and Applied Linguistic, University of Birmingham, Birmingham, United Kingdom,

² Department of Applied Sciences, University of the West of England, Bristol, United Kingdom, ³ Language and Cognition Department, Max Planck Institute of Psycholinguistics, Nijmegen, Netherlands, ⁴ School of Psychology, University of Birmingham, Birmingham, United Kingdom

OPEN ACCESS

Edited by:

Wendy Sandler,
University of Haifa, Israel

Reviewed by:

Francesca Peressotti,
Università degli Studi di Padova, Italy
Laura J. Speed,
Radboud University Nijmegen,
Netherlands

*Correspondence:

Marcus Perlman
m.perlman@bham.ac.uk

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 15 December 2017

Accepted: 23 July 2018

Published: 14 August 2018

Citation:

Perlman M, Little H, Thompson B and
Thompson RL (2018) Iconicity
in Signed and Spoken Vocabulary:
A Comparison Between American
Sign Language, British Sign
Language, English, and Spanish.
Front. Psychol. 9:1433.
doi: 10.3389/fpsyg.2018.01433

Considerable evidence now shows that all languages, signed and spoken, exhibit a significant amount of iconicity. We examined how the visual-gestural modality of signed languages facilitates iconicity for different kinds of lexical meanings compared to the auditory-vocal modality of spoken languages. We used iconicity ratings of hundreds of signs and words to compare iconicity across the vocabularies of two signed languages – American Sign Language and British Sign Language, and two spoken languages – English and Spanish. We examined (1) the correlation in iconicity ratings between the languages; (2) the relationship between iconicity and an array of semantic variables (ratings of concreteness, sensory experience, imageability, perceptual strength of vision, audition, touch, smell and taste); (3) how iconicity varies between broad lexical classes (nouns, verbs, adjectives, grammatical words and adverbs); and (4) between more specific semantic categories (e.g., manual actions, clothes, colors). The results show several notable patterns that characterize how iconicity is spread across the four vocabularies. There were significant correlations in the iconicity ratings between the four languages, including English with ASL, BSL, and Spanish. The highest correlation was between ASL and BSL, suggesting iconicity may be more transparent in signs than words. In each language, iconicity was distributed according to the semantic variables in ways that reflect the semiotic affordances of the modality (e.g., more concrete meanings more iconic in signs, not words; more auditory meanings more iconic in words, not signs; more tactile meanings more iconic in both signs and words). Analysis of the 220 meanings with ratings in all four languages further showed characteristic patterns of iconicity across broad and specific semantic domains, including those that distinguished between signed and spoken languages (e.g., verbs more iconic in ASL, BSL, and English, but not Spanish; manual actions especially iconic in ASL and BSL; adjectives more iconic in English and Spanish; color words especially low in iconicity in ASL and BSL). These findings provide the first quantitative account of how iconicity is spread across the lexicons of signed languages in comparison to spoken languages.

Keywords: sign language, spoken language, iconicity, modality, American Sign Language, British Sign Language, English, Spanish

INTRODUCTION

Increasingly, language scientists recognize that iconicity – in complement to arbitrariness – is a fundamental feature of human languages (Perniss et al., 2010). On this theory, many of the forms of languages, from phonology to morphology to syntax, are motivated by a resemblance to the meaning they are used to express. Recent studies indicate that iconicity can serve several important functions, playing a role in how language is produced and processed, how it is learned and acquired by children, how it changes over history, and indeed, how it evolved in the first place (see reviews in Imai and Kita, 2014; Perlman and Cain, 2014; Perniss and Vigliocco, 2014; Dingemanse et al., 2015; Meir and Tkachman, 2018).

Building on the growing documentation of iconic phenomena across languages both signed and spoken, including the spontaneous gestures that are integrated with sign and speech, researchers are beginning to apply a comparative perspective to the study of iconicity (e.g., Voeltz and Kilian-Hatz, 2001; Kita and Özyürek, 2003; Dingemanse, 2012; Padden et al., 2013; Perry et al., 2015; Östling et al., 2018). Although most previous research has focused separately on signed or spoken languages, a comparative approach raises fundamental questions related to the modality of language. How is iconicity manifested in languages that are signed, compared to those that are spoken? Is it true that signed languages are categorically more iconic than spoken languages, as it is often assumed? Or might there be more interesting, richer differences – as well as similarities – in the patterns of iconicity found in signed and spoken languages?

In this paper, we examine whether the visual-gestural modality of signed languages facilitates iconicity for different kinds of lexical meanings than the auditory-vocal modality of spoken languages. Our study analyzes iconicity in the vocabularies of two signed languages – American Sign Language (ASL) and British Sign Language (BSL), and two spoken languages – English and Spanish. Using previously collected iconicity ratings of signs and words¹, we directly compared how semantics motivates iconicity across the lexicons of the four different languages.

Iconicity in Signed and Spoken Languages

It is widely taken for granted that signed languages are categorically more iconic than spoken languages. Many scholars have observed that signed languages, which are based on visible actions of the hands, body and face, are well suited for iconic representation (Johnston and Schembri, 1999; Meier, 2002; Meir, 2010; Cartmill et al., 2012; Kendon, 2014). Stemming from this potential, the iconicity in signed languages is widespread and clearly evident in both their grammar and, of focus here, their lexicon (Klima and Bellugi, 1979; Armstrong et al., 1995; Taub, 2001; Liddell, 2002; Aronoff et al., 2005). For example, Stokoe (1965) identified 25% of ASL signs to be either pantomimic or iconic, and Wescott (1971) further estimated that of the remaining 75% of signs, about two-thirds seemed plausibly

derived from iconic origins. A more rigorous analysis examining 1944 signs of Italian Sign Language found that 50% of handshapes and 67% of body locations appeared to have iconic motivation (Pietrandrea, 2002). Indeed, many signs are iconic to such a degree that it was not until the pioneering work of Stokoe (1965) that they were even recognized by linguists as the components of legitimate languages, rather than idiosyncratic pantomimes and “mimic” signs (Wilcox, 2004).

Recent studies show that iconicity in the vocabularies of signed languages is not exhibited haphazardly. Signs for some kinds of meanings tend to be more iconic than others. For example, an analysis of BSL found that signs for objects and actions were more iconic than signs for properties (Perniss et al., 2017), presumably because manual signals afford more iconicity for objects and actions. Some patterns of iconicity have been shown to be common across a large number of signed languages. Lepic et al. (2016) found, across four unrelated languages, that two-handed signs are more frequently associated with meanings that are inherently “plural.” A larger-scale study based on the automated visual processing of signs from 31 different languages similarly found a correlation between the use of two-handed forms for signs and the degree of plurality in their meaning (Östling et al., 2018). Östling et al. (2018) also analyzed signs with sensory and body part-related meanings, where they found a correlation between the anatomical meaning of a sign and the location on the signer’s body where it is articulated.

In addition, comparative studies – such as by Meir et al. (2013) – have observed that patterns of iconicity in the lexicon of a signed language can vary systematically between languages. For example, Padden et al. (2013) examined the iconic strategies that signers used to represent hand-held tools (e.g., comb, mop, handsaw) in ASL, New Zealand Sign Language (NZSL), and Al-Sayyid Bedouin Sign Language (ABSL). Their analysis compared the use of three primary iconic strategies used by signers to represent the objects: signing as if handling the object (handling); signing as if using the object, but with the hands shaped to display qualities of its shape (instrument); and signing as if the hands are the object, but without performing its characteristic action (object). The results showed that, compared to hearing non-signers, signers of all three languages more strongly preferred the instrument strategy over the handling strategy. Notably, the signers of different languages also showed different proclivities: signers of ASL and ABSL displayed a stronger preference for the instrument strategy than NZSL signers.

In another comparative study, Meir et al. (2013) examined how signers use their body as a resource for the iconic representation of actions involving particular body parts. The authors observed that signers can make use of two different iconic strategies with respect to indicating the participants of an action. They can use their body as the subject of an action (e.g., the signer represents the subject of ‘eat’, i.e., the eater), without indicating the person of the participant. Or they can use their body to indicate a first person participant in an action in opposition to directing the sign toward locations in space associated with non-first person participants (representing ‘I eat’ vs. ‘you eat’). Meir et al. (2013) then compared the strategies used in the signs of ABSL, a young signed language, to those of

¹ We take the convention of referring to lexical items in signed languages as *signs*, and to lexical items in spoken languages as *words*.

Israeli Sign Language (ISL), a more mature language. For ISL, signs were elicited from three different age groups of signers, providing diachronic perspective on the language. The study found that ABSL signers only used signs implementing the body-as-subject strategy, without encoding person distinctions – a pattern that was also predominant with older ISL signers. In contrast, younger ISL signers – representing the more mature stage of the language – made use of a person agreement strategy in which they directed some signs toward locations in space to show who did what to whom. Thus, the body-as-subject strategy appears to be more basic and prevalent across signed languages, whereas the agreement strategy within ISL may be adopted by more mature languages in which a lexical class of agreement verbs is created over time.

In contrast to signed languages, the predominant linguistic theories over history have widely assumed that the vocabularies of spoken languages are essentially arbitrary (e.g., de Saussure, 1959; Hockett, 1960; Pinker and Bloom, 1990). A common line of explanation for this builds on the assumption that the auditory-vocal modality affords far less potential for iconicity than the visual-gestural modality (e.g., Hockett, 1978; Armstrong and Wilcox, 2007; Tomasello, 2008; Meir et al., 2013). The clear exception of onomatopoeia – the iconic representation of sounds – has often been trivialized, without much attempt at rigorous empirical justification. For example, Newmeyer (1992), citing Whitney (1874), referred to “the almost complete non-existence of an iconic relationship between words and their referents,” suggesting that “the number of pictorial, imitative, or onomatopoeic non-derived words in any language is vanishingly small” (p. 758)². Similarly, Pinker (1999) observed that “onomatopoeia and sound symbolism certainly exist, but they are asterisks to the far more important principle of the arbitrary sign” (p. 2). And in this vein, a popular introduction to psycholinguistics acknowledges that onomatopoeic words such as *cuckoo*, *pop*, *bang*, *slurp*, and *squish* are exceptions to the principle of arbitrariness, but observes that “there are relatively few of these in any language” (Aitchison, 2007, p. 29).

Nevertheless, over the years, researchers have collected wide-ranging evidence of iconicity in the vocabularies of spoken languages (Perniss et al., 2010; Perlman and Cain, 2014; Dingemanse et al., 2015). As a baseline, it turns out that onomatopoeic words are much more prevalent than the above quotes suggest, and indeed, may even constitute a distinct lexical class that is universal across languages (Dingemanse, 2012). For example, although English has been characterized as a spoken language with a vocabulary that is relatively lacking in iconicity (e.g., Vigliocco et al., 2014), studies of onomatopoeic words in English reveal a substantial and varied inventory. For instance, an analysis by Rhodes (1994) examined over one hundred English words used to refer to “aural images,” including predominantly onomatopoeic words. These words spanned diverse conceptual categories, including sounds produced by the vocal tracts of

humans (e.g., *yell*, *hum*), and other animals (e.g., *moo*, *tweet*), as well as non-vocal sounds (e.g., *click*, *bang*). While somewhat narrower in scope, the use of iconic words to represent vocal tract actions in spoken languages can be seen as an analog to the iconic representation of various kinds of manual actions in signed languages.

Rhodes (1994) noted that onomatopoeic words fall broadly along a continuum in the degree to which they are conventionalized into the lexicon (also see Dingemanse and Akita, 2017). On one end are tame words, which are highly lexicalized and characterized by standard phonological and syntactic patterns. On the farthest end of tame, Rhodes observed that a few aural images are conveyed by standard arbitrary words such as *noise*, *sound*, and *din*. On the other end of the scale are wild words, which utilize the full range of the vocal tract to precisely imitate sounds (also see Lemaitre et al., 2016). Both the prevalence and the productiveness of onomatopoeia in English are illustrated in dictionaries, such as the online *Written Sound Onomatopoeia Dictionary*, which contains 772 entries³ (retrieved 1/11/2017), and *KA-BOOM!* – a dictionary of onomatopoeia in comic books – which contains 119 pages and thousands of entries (Taylor, 2007). The quickly evolving contents of these dictionaries pay tribute to the dynamic quality of onomatopoeia, which can serve as a productive source of lexical innovation, perhaps comparable in ways to the creative functions of iconicity in signed languages (cf. Klima and Bellugi, 1979).

Beyond the often-underestimated base of onomatopoeia, a growing number of cross-linguistic studies show that iconicity in spoken languages is far from limited to the representation of sounds. In most languages around the world, onomatopoeic words typically represent just a portion of a semantically broader class of vocabulary – variously termed mimetics, expressives, phonaesthemes, and most generically, *ideophones* – that are used to communicate about an array of concepts related to the senses (Diffloth, 1972; Nuckolls, 1999; Voeltz and Kilian-Hatz, 2001; Dingemanse, 2012; Kwon, 2015). As a general class, ideophones are characterized as marked words that are used to convey sensory imagery (Dingemanse, 2012). They are noted for their special forms and distinct grammatical properties, e.g., the use of reduplication as an iconic representation of repetition. Ideophones subsume onomatopoeia, and they broadly comprise a dynamic class of words that is commonly associated with creativity and lexical innovation (Dingemanse, 2014). In addition to sound, ideophones are used to express meanings from varied semantic domains such as luminance, manner of movement, size, texture, shape, taste, temperature, and emotional and psychological states (Dingemanse, 2012). This range is illustrated by Dingemanse (2012, p. 661), which presented an assortment of examples from seven diverse languages: for example, *gùdùù* ‘pitch dark’ (Siwu), *juluq* ‘to gulp down (something solid) without chewing’ (Somali), *dzing* ‘a sudden awareness or intuition, especially one that causes fright’ (Pastaza Quechua), *potil* ‘soft and tender (surface)’ (Korean), *kiláv-kiláv* ‘in a zigzagging motion’ (Ngbaka Gbaya), *lipilip* ‘sparkling like a diamond or piece of glass’

²Citing Klima and Bellugi (1979), Newmeyer also claimed that, “We now know that even the signed languages of the Deaf, where one might intuitively expect it, manifest little sign-referent iconicity” (p. 758). This characterization of Klima and Bellugi’s study of iconicity in ASL appears to miss the considerable complexity and nuance in the findings of their research and their conclusions.

³www.writtensound.com

(Upper Necaxa Totonac), and *blbʔə* ‘painful embarrassment’ (Semai).

Native speakers typically have the impression that ideophones are distinctly depictive and often iconic of their meaning, and comprehension experiments with naïve listeners provide some support for these intuitions (Kantartzis et al., 2011; Dingemanse et al., 2016; also see Kwon, 2017 for English phonaesthemes). For example, Dingemanse et al. (2016) tested the ability of naïve listeners to understand the meanings of ideophones from a diverse sample of unfamiliar languages. The stimuli represented five different semantic domains, including color/vision, motion, shape, sound, and texture. Although listeners were most accurate at guessing the meanings of ideophones for sound concepts, their guessing was significantly above chance for each domain.

In light of cross-linguistic surveys indicating the widespread prevalence of ideophones, and especially their semantic diversity, some linguists have proposed that studies of ideophones call for typological and comparative approaches (Diffloth, 1972; Voeltz and Kilian-Hatz, 2001; Dingemanse, 2012, 2017a). In this spirit, Dingemanse (2012, p. 663) proposed an implicational hierarchy for the semantic range of ideophone systems across languages. According to this hierarchy, almost all spoken languages have ideophones for sound concepts. As an ideophone system becomes richer and more varied, it tends to expand first to encompass meanings related to movement, then visual patterns and other sensory perceptions, and finally inner feelings and cognitive states. The prominence of sound in this semantic hierarchy may correspond to the higher potential for iconicity in sound-related ideophones (cf. Dingemanse et al., 2016).

In addition to ideophones (including onomatopoeia), a number of cross-linguistic studies show that some words outside of this distinct class may also be iconic, especially within a few particular semantic domains. For example, across many languages, words that express the concept of ‘small’ are more likely to have higher-frequency vowels, such as the high front vowel in the English *teeny*, compared to the low-frequency back vowel in *huge* (Ultan, 1978; Ohala, 1994; Haynie et al., 2014; Blasi et al., 2016). This may reflect the tendency for smaller things – particularly vocalizing animals – to emit higher-frequency sounds, compared to larger things, which tend to emit sounds of lower frequency. A similar pattern is reflected in male and female personal names, e.g., *Emily* and *Thomas* (Pitcher et al., 2013), and in indexical words used to refer to proximal and distal referents, e.g., *this* and *that*, *near* and *far* (Tanz, 1971; Ultan, 1978; Johansson and Zlatev, 2013).

Similar to signed languages in which the body serves as an iconic naming device for anatomical meanings (Meir et al., 2013), in spoken languages, we find an analog in the articulation of words used to refer to parts of the vocal tract (Wichmann et al., 2010; Urban, 2011; Blasi et al., 2016). For example, evidence from statistical studies across large, diverse samples of languages indicates that words for ‘lip’ tend to feature bilabial consonants (as do words for ‘breast,’ perhaps related to suckling). In addition, words for ‘nose’ tend to feature nasal phonemes, and words for ‘tongue’ the lateral /l/.

Considered together, these studies illustrate how iconicity is a prevalent characteristic of signed and spoken lexicons alike. Crucially, iconicity appears to be spread systematically across the semantic space of a language in ways that correspond with the iconic resources of its modality. For example, in signed languages, iconicity is high in words related to (non-vocal) bodily actions, whereas in spoken languages iconicity is more concentrated in words related to vocal tract actions and sounds. Indeed, Dingemanse et al. (2015) has postulated some basic commonalities and differences that might characterize iconicity in the lexicons of signed versus spoken languages based on the semiotic affordances of each modality. They suggested that meanings related to qualities like ‘size,’ ‘repetition,’ ‘temporal unfolding,’ and ‘intensity’ may readily lend themselves to iconicity in both modalities. Meanings related to ‘spatial relations’ and ‘visual shape’ may afford iconicity in signed languages, but not spoken ones, while ‘sound’ and ‘loudness’ may afford considerable iconicity in spoken words, but not signs. Semantic domains like ‘abstract concepts’ and ‘logical operators’ may be hard for both types of languages to represent with iconic forms.

Iconicity Ratings of Signs and Words

While cross-linguistic studies provide suggestive evidence for hypotheses such as those of Dingemanse et al. (2015), a more decisive investigation requires broader, systematic analyses of how iconicity is spread across the lexicons of individual languages. To this end, some recent studies have used iconicity ratings collected for large numbers of signs (e.g., Vinson et al., 2008; Caselli et al., 2017) and words (e.g., Perry et al., 2015; Winter et al., 2017a). For example, an original study of BSL collected iconicity ratings for 300 signs sampled from various sources to include a range of iconic and non-iconic signs (Vinson et al., 2008). The signs were rated by deaf signers on a scale from 1 (not iconic) to 7 (most iconic). The results showed that the iconicity of signs was negatively correlated with the age at which they are typically acquired: signs learned earlier tended to be more iconic (also see Thompson et al., 2012 for similar findings with children). There was also a small positive correlation between iconicity and the familiarity of signs. Iconicity ratings of a larger, more widely representative sample of ASL signs – these rated by hearing non-signers – found that signs were skewed toward the arbitrary end of the 1-to-7 scale (Caselli et al., 2017). Iconicity ratings showed a weak negative correlation with frequency, and a positive correlation with neighborhood density – that is, more iconic signs tended to be similar in form to more signs. A follow-up study with a subset of the ASL signs found a similar relationship between iconicity and age of acquisition to that of BSL (Caselli and Pyers, 2017). Thus, the use of iconicity ratings revealed interesting patterns in the distribution of iconicity across the lexicons of these two, unrelated signed languages.

Recently, a series of studies applied a similar approach to study iconicity in the vocabularies of spoken languages. The first of these studies compared iconicity in roughly 600 words in English and in Spanish (Perry et al., 2015). Notably, English and Spanish are Indo-European languages, which – it has been claimed – are less iconic than most other spoken languages (Perniss et al., 2010).

This idea is illustrated by Vigliocco et al. (2014, p. 2): “Indeed, if we look at the lexicon of English (or that of other Indo-European languages), the idea that the relationship between a given word and its referent is defined by an arbitrary connection alone seems entirely reasonable. For example, there is nothing in the sequence of sounds in the English word ‘house’ that indicates its meaning of ‘a building for human habitation’.” However, contrary to this line of reasoning, the results of Perry et al. (2015) demonstrate that the vocabularies of English and Spanish are iconic in measurable, theoretically interesting ways. For example, in both languages, as in BSL and ASL, the iconicity ratings of words were negatively correlated with their age of acquisition – even when excluding onomatopoeia (also see Massaro and Perlman, 2017; Perry et al., 2017). Thus, it appears that young English and Spanish speaking children are sensitive to the iconicity of words, and they pick up on the more iconic words first.

The ratings also gave an opportunity to examine whether iconicity in languages like English and Spanish – which lack rich ideophone systems – might nevertheless pattern according to certain semantic dimensions, such as those postulated by Dingemanse (2012) and Dingemanse et al. (2015). Indeed, when iconicity was compared between lexical classes, some noteworthy patterns emerged. In English, Perry et al. (2015) found that onomatopoeia and interjections were highest in iconicity, followed by adjectives and verbs, then nouns, and finally closed-class function words. This pattern roughly corresponds with the ordering of Dingemanse’s (2012) implicational hierarchy, which proposed that ideophones are most prevalent for the expression of sound concepts, followed by concepts related to motion, vision, and other sensory perceptions. Similarly, Imai and Kita (2014) noted that ideophones typically have a rich inventory for expressing manners of action, physical sensations and certain properties of objects, but are not often used to refer directly to objects. Thus, it fits that in English, onomatopoeia, and then verbs – typically relating to motion and action, and adjectives – relating to sensations and properties, would be most iconic. Likewise, it fits that nouns, which often refer to objects, would be relatively low in iconicity. Furthermore, the low ratings for function words may reflect Dingemanse et al.’s (2015) suggestion that logical relations are not amenable to iconic representations.

The results for Spanish were comparable to those of English, but with one key difference that may stem from a typological difference between the languages. Perry et al. (2015) noted that English and Spanish vary in the typology of their verbs (Talmy, 2000; Beavers et al., 2010). English is a satellite-framed language, which typically conveys manner of motion in the main verb. In contrast, Spanish is a verb-framed language – verbs tend to convey the path of motion, and leave information about the manner for expression in adverbials. For example, consider the English sentence “The bottle floated into the cave,” in which manner of movement is expressed by the verb. Compare this to the Spanish “La botella entró a la cueva flotando,” in which the manner of the action, “flotando” (floating) is separated from the main verb “entró” (moved-in). Because Spanish verbs tend not to express manner of motion, a rich source of iconicity in many ideophones, Perry et al. (2015) predicted that these words would be less iconic. In line with this hypothesis, the results showed that

the iconicity of Spanish verbs was low compared to adjectives, and more comparable to nouns and function words.

A subsequent study with English expanded on Perry et al.’s (2015) iconicity ratings to include a total of 3001 words (Winter et al., 2017a). This study found essentially the same pattern of results with respect to lexical class: onomatopoeia and interjections were highest in iconicity, then verbs, adjectives, nouns, and finally function words. Winter et al. further examined the specific semantic factors that might influence the distribution of iconicity across English vocabulary. First, they tested whether the iconicity ratings correlated with ratings of the degree to which a word “evokes a sensory experience” (Juhasz and Yap, 2013). In a model that also included imageability ratings and frequency, sensory experience was the strongest predictor of iconicity. As with ASL (Caselli et al., 2017), frequency was negatively correlated with iconicity. Imageability was also negatively correlated with iconicity, suggesting that more highly visual words may be less iconic. A subsequent analysis categorized the meanings of the words into their dominant sensory modality. This showed that words with meanings most strongly associated with the auditory and the tactile modalities were rated higher in iconicity than those associated with the other modalities, with visual words even lower than olfactory and gustatory words.

Another set of analyses by Sidhu and Pexman (2017) with the English iconicity ratings found a similar relationship between iconicity and the sensory experience ratings of Juhasz and Yap (2013). Additionally, this study found that the strongest relationship was between iconicity and semantic neighborhood density, which mediated the effect of sensory experience. Words in sparser semantic neighborhoods, particularly those high in sensory experience, tended to be more iconic, a result that held across adjectives and adverbs, verbs, and nouns. This finding supports the idea that lexicons exhibit a balance between iconicity and arbitrariness: as more words share similar meanings, the ability to discriminate between them becomes more critical, which drives them toward more arbitrary forms (Gasser, 2004; Dingemanse et al., 2015).

In summary, these studies with iconicity ratings show some of the various ways that iconicity is systematically distributed across the lexicons of signed and spoken languages. Some of these patterns appear to be common to both language modalities. For example, there is a consistent relationship between the iconicity of a sign or word and the age at which it is learned by children: more iconic items tend to be learned earlier. Yet, these studies also hint at some notable differences between the iconicity in signed and spoken vocabularies, particularly with respect to different semantic domains. For instance, in spoken languages, adjectives and particularly words for auditory and tactile properties appear to be relatively high in iconicity, whereas this may not be the case in signed languages (cf. Perniss et al., 2017). Conversely, nouns and visual words in spoken languages appear to be low in iconicity, while there is reason to think they are more highly iconic in signed languages.

Current Study

In the current study, we conducted a direct comparison between the iconicity of signed and spoken vocabularies and how it

varies across different semantic domains. We asked whether the gestural-visual modality of signed languages motivates iconicity for different kinds of meanings than the vocal-auditory modality of spoken languages. To investigate this question, we utilized previously collected iconicity ratings to compare ASL and BSL with English and Spanish. Across the four languages, we examined: (1) the correlations between languages for iconicity ratings of the same meanings; (2) the relationship between iconicity ratings and an array of ratings for various semantic properties (e.g., concreteness, sensory experience); (3) how iconicity ratings vary broadly between (English-based) lexical classes; and (4) how they vary between more specific semantic categories (e.g., clothes, colors).

MATERIALS AND METHODS

Iconicity Ratings

Our study utilized previously collected iconicity ratings for 100s of signs and words in ASL, BSL, English, and Spanish. These samples of signed and spoken languages were chosen opportunistically because of the pre-existing data available, and not because ASL and BSL share any particular comparable relationship to English and Spanish. Notably, ASL and BSL are not historically related to each other, whereas English and Spanish share common ancestry as Indo-European languages. Moreover, there is even more recent shared ancestry between the spoken languages as English contains a large amount of Latinate vocabulary from French.

Table 1 shows information about the source of the ratings for each language, the participants who provided the ratings, and the number of signs and words covered. Iconicity ratings for 993 ASL signs come from Caselli et al. (2017; see the LEX-ASL database⁴). The signs were rated by non-signers on a scale from 1 (“not iconic at all”) to 7 (“extremely iconic”). Iconicity ratings for 604 BSL signs are from Vinson et al. (2008) and Thompson et al. (unpublished). In these studies, signs were rated 1 (arbitrary) to 7 (iconic) by a mix of native and non-signers in four different experiments. When ratings for a given BSL sign were collected in multiple experiments, we used an averaged rating for our analysis. We found that our different ratings for BSL signs were highly correlated, including between signers and non-signers ($r \geq 0.84$ for all sets of overlapping ratings). This was comparable to a study of ASL with a different set of signs, which found a correlation of $r = 0.82$ between the ratings of ASL signs by signers and non-signers (Sevcikova Sehyr et al., 2017).

⁴<http://www.asl-lex.org/>

Iconicity ratings for 3001 English words were collected by Perry et al. (2015) and Winter et al. (2017a). Native speakers rated the words on a scale from -5 (sounds like the opposite of what it means) to 5 (sounds like what it means), with 0 (arbitrary) at the middle point. Iconicity ratings for 637 Spanish words come from Perry et al. (2015). These were provided by native speakers and collected according to a similar procedure as the English ratings. In a few instances, multiple Spanish words shared the same English gloss, e.g., Spanish *un* and *una* translate to English *a*, and Spanish *poco* and *poquito* to English *little*. For these cases, we selected the variant with the higher iconicity rating for our analyses.

We direct the reader to the original sources for further information on the procedures used to collect the ratings, including the particular instructions and examples used to define ‘iconicity.’ One detail of the instructions that is worth noting here is that they all reflected the modality of the language: for signs, iconicity was defined as when a sign ‘looks’ like what it means, and for spoken words, as when a word ‘sounds’ like what it means.

Ratings for Semantic Properties

We investigated the relationship between the iconicity ratings and a battery of ratings related to various semantic properties of words: concreteness, imageability, sensory experience, and perceptual strength for vision, audition, touch, gustation, and olfaction. **Figure 1** shows the number of items in each language for which we had each measure. Notably, each of these measures was collected with respect to English words, and thus, our analyses of words and signs of other languages use the ratings from their English translations.

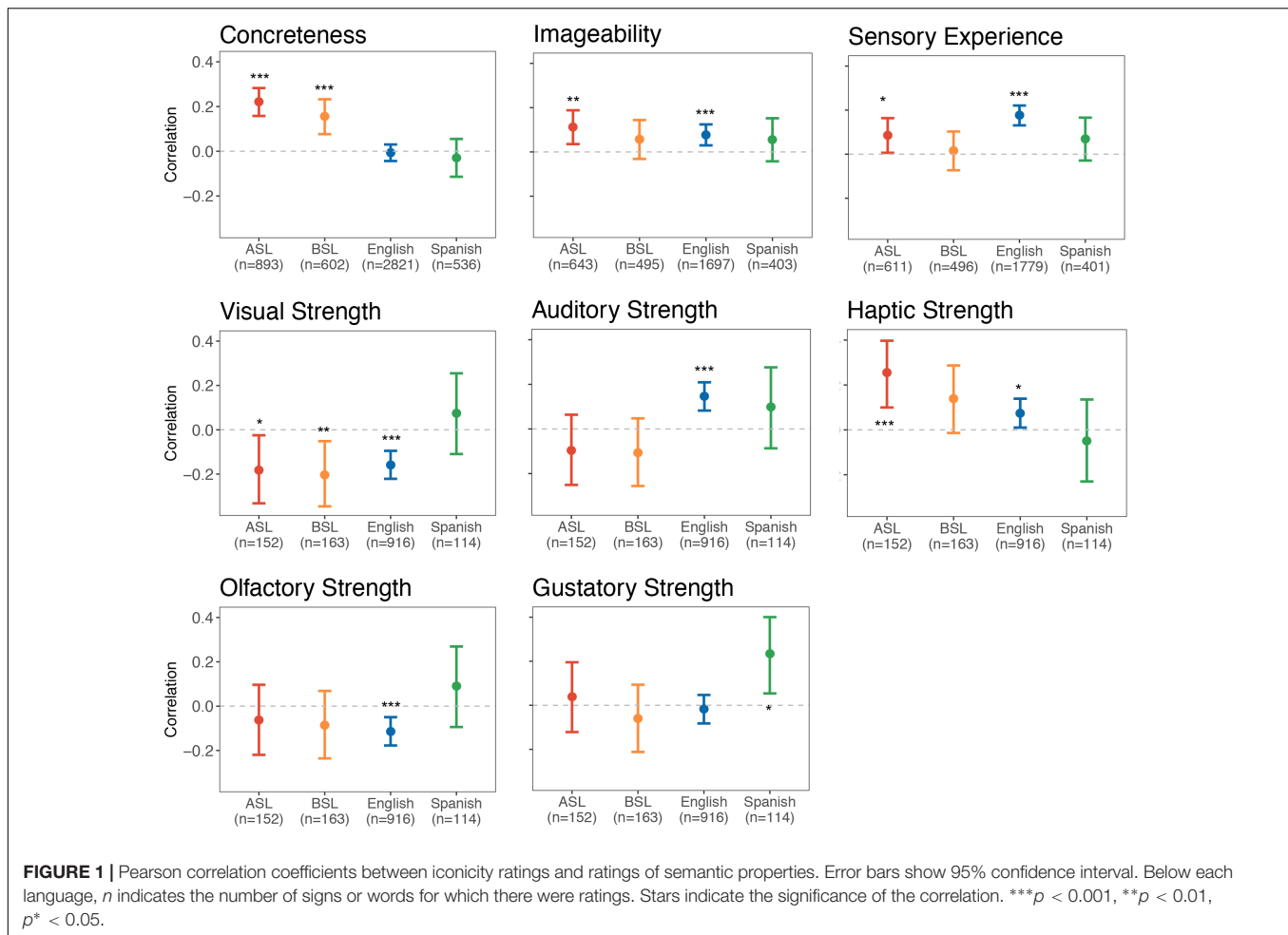
Concreteness ratings come from Brysbaert et al. (2014). Participants were asked to rate the degree to which the referent of a word was experienced “directly through one of the five senses.” Imageability ratings come from Cortese and Fugett (2004), which assessed how much the meanings of words related to “sensory experience, such as a mental picture of sound.” Sensory experience ratings were collected by Juhasz and Yap (2013). For these ratings, participants rated the degree to which a word evoked a sensory experience. Finally, perceptual strength ratings include ratings of verbs from Winter (2016), adjectives from Lynott and Connell (2009), and nouns from Lynott and Connell (2013). These ratings measured the degree to which a word was associated with each of the five sensory modalities.

Lexical Classes

Our analyses of iconicity and lexical class focused on the 220 meanings for which we had iconicity ratings in all

TABLE 1 | Source of iconicity ratings for each language.

Modality	Language	No. signs	Raters	References
Signed	ASL	993	Non-signers	Caselli et al., 2017
	BSL	604	Mix of native/non-signers	Vinson et al., 2008; Thompson et al., unpublished
Spoken	English	3001	Native speakers	Perry et al., 2015; Winter et al., 2017a
	Spanish	637	Native speakers	Perry et al., 2015



four languages. To sort the meanings into broad semantic categories, we assigned each one to a lexical class based on its part of speech in English, adapted from Brysbaert and Keuleers' (2012) annotation of the SUBTLEX corpus. This resulted in 132 nouns, 41 verbs, 28 adjectives, and 19 grammatical words and adverbs (not typically related to manner). **Table 2** shows the lexical class for each of these words.

The use of lexical class to classify meanings into broad semantic categories is supported by theories of cognitive grammar, which posit that classes such as nouns, verbs, and adjectives reflect conceptual prototypes (Givón, 2001; Langacker, 2008; also see Strik Lievers and Winter, 2018). For example, according to Langacker, nouns are rooted in the prototype of a physical object, a 'thing' (i.e., subsuming people and places and not limited to physical entities); verbs typically refer to actions and events, profiling change over time; and adjectives typically specify more static properties. Nevertheless, we emphasize that our designated lexical categories, based on English, serve just for a broad-stroke comparison of how different kinds of meanings vary in iconicity between the languages. We do not mean to imply that ASL, BSL, and Spanish necessarily share these same lexical classes with English.

Specific Semantic Categories

We further classified each of the 220 meanings into more specific semantic categories. As shown in **Table 2**, these included nine classes of nouns, three classes of verbs, four classes of adjectives, and four classes of grammatical words and adverbs. These specific categories were determined *ad hoc* from the sample of 220 meanings that happened to feature ratings in each language. Accordingly, their purpose is to present a more detailed – but exploratory – breakdown of iconicity across the four vocabularies.

Data Availability

Data and analysis scripts are made available through the Open Science framework at <https://osf.io/d759h/>.

RESULTS AND DISCUSSION

Correlation of Iconicity Between Languages

First, we calculated the correlation between the iconicity ratings of each pair of languages. The correlation between ASL and BSL signs was fairly strong, $r = 0.68$, $t(344) = 17.0$, $p < 0.0001$.

TABLE 2 | Lexical class and particular semantic categories of the 220 meanings for which we had iconicity ratings in all four languages.

Lexical class	Semantic category	Meanings
Nouns	Small artifacts (roughly fits within two hands)	Ball, balloon, book, bottle, doll, hammer, knife, medicine, money, napkin, paper, pencil, scissors, soap, spoon, telephone
	Body parts and clothes	Arm, dress, ear, glasses, hair, hat, mouth, necklace, scarf, shirt, shorts, shoulder, skirt
	Vehicles	Airplane, bicycle, boat, car, helicopter, motorcycle, train
	Food	Apple, banana, bread, butter, cake, cereal, cheese, chocolate, coffee, egg, hamburger, meat, milk, potato, salt, soup, spaghetti, strawberry
	Animals	Animal, bear, bird, butterfly, cat, cow, crocodile, deer, dog, duck, elephant, giraffe, horse, lion, monkey, mouse, owl, pig, rabbit, sheep, snake, squirrel, tiger, turkey, turtle, wolf
	Natural things	Cloud, fire, flower, grass, moon, rain, star, sun, tree, water
	Buildings and rooms	Bedroom, church, garage, house, kitchen, school
	People	Baby, boy, brother, clown, doctor, family, friend, girl, man, nurse, person, sister, uncle, woman
	Other things (bigger and/or more abstract things)	Bed, blanket, box, chair, clock, computer, country, day, door, dryer, flag, mirror, morning, movie, night, picture, pillow, shower, stairs, street, table, window
	Colors	Black, blue, green, orange, red, yellow
Adjectives	Evaluations	Bad, cute, good, poor, right
	Feelings and emotions	Cold, hungry, sad, scared, sick, tired
	Other properties	Dirty, full, hard, new, old, quiet, same, slow, small, tall, wet
Verbs	Manual actions	Brush, draw, drink, eat, paint, pull, push, sweep, throw, tie, write
	Locomotion	Come, fall, fly, go, jump, run, skate, swing, walk
	Other actions	Break, buy, cook, cry, find, finish, have, help, make, play, read, see, show, sing, sit, sleep, talk, think, wait, want, win
Adverbs and grammatical words	Spatial and determiners	Behind, here, more, some, that
	Time	After, later, now, yesterday
	Question words	How, what, when, where, which, who
	Other grammatical words	And, but, you, with

English showed a small but reliable correlation with ASL, $r = 0.16$, $t(550) = 3.7$, $p < 0.0001$, BSL, $r = 0.22$, $t(601) = 5.4$, $p = 0.0003$, and Spanish, $r = 0.16$, $t(478) = 3.6$, $p = 0.0003$. Spanish ratings did not significantly correlate with BSL, $t(323) = 0.2$, $p = 0.83$, and showed a weak, negative correlation with ASL, $r = -0.12$, $t(275) = -2.02$, $p = 0.04$.

These results suggest that signs for particular meanings are fairly consistent in their level of iconicity in ASL and BSL, while there is greater variability between English and Spanish words. This pattern may reflect that potential iconic mappings between form and meaning are more direct and transparent for many signs, and hence more consistently realized across different signed languages. In comparison, words may reflect vaguer, less obvious iconic mappings between form and meaning, which, as a consequence, appear less consistently across spoken languages.

Intriguingly, in addition to being correlated with Spanish, the iconicity of English words was also weakly, but positively, correlated with the iconicity of the corresponding ASL and BSL signs. However, this was not the case with Spanish, which showed – if anything – a negative correlation with the signed languages. In part, this may stem from the low iconicity of verbs in Spanish in comparison to English, as was previously reported by Perry et al. (2015). Accordingly, English may share with ASL and BSL relatively high iconicity in verbs, but shares other features of iconic vocabulary with Spanish. The following analyses examine how iconicity is spread across

these four vocabularies in more detail, shedding light on their commonalities and differences.

Iconicity and Semantic Properties

For each language, we examined the relationship between iconicity ratings and ratings of a host of semantic properties: concreteness, imageability, sensory experience, and perceptual strength with respect to vision, audition, touch, gustation, and olfaction. **Figure 1** shows plots of the correlations between the iconicity ratings in each language and these variables. To test whether the strength of these relationships differed between language modalities (i.e., signed or spoken), we constructed linear mixed-effects models with the ratings of each semantic property as a predictor of iconicity ratings. The models included main effects for the semantic variable and modality (both centered), and a term for their interaction. Random intercepts were included for language and meaning, and random slopes were included for the semantic variable on language. Significance tests were calculated using χ^2 -tests that compared the model likelihoods with and without the factor of interest.

The model for concreteness ratings showed that concreteness was a significant predictor of iconicity, $b = 0.14$, 95% CI = [0.08, 0.20], $\chi^2_1 = 4.42$, $p < 0.01$. More concrete meanings tended to have more iconic signs and words. There was also a significant interaction between concreteness and modality, $b = -0.28$, 95% CI = [-0.40, -0.17], $\chi^2_1 = 9.04$, $p < 0.01$. This indicated that

concreteness was more highly correlated with iconicity ratings in signed languages.

The model for sensory experience ratings showed that sensory experience was a significant predictor of iconicity ratings, $b = 0.13$, 95% CI = [0.05, 0.21], $\chi^2_1 = 7.17$, $p < 0.01$. Meanings higher in sensory experience were associated with more iconic signs and words. There was not a significant interaction between modality and sensory experience, $\chi^2_1 = 0.96$, n.s.

The model for imageability ratings (scaled to z-scores) showed that imageability was a significant predictor of iconicity ratings, $b = 0.11$, 95% CI = [0.04, 0.17], $\chi^2_1 = 6.14$, $p < 0.05$. More imageable meanings tended to have more iconic signs and words. There was not a significant interaction between imageability and modality, $\chi^2_1 = 1.95$, n.s.

The model for visual strength ratings showed that visual strength was a significant predictor of iconicity ratings, $b = -0.18$, 95% CI = [-0.29, -0.06], $\chi^2_1 = 6.92$, $p < 0.01$. Meanings with greater visual strength tended to have less iconic signs and words. There was also a significant interaction between visual strength and modality, $b = 0.20$, 95% CI = [0.01, 0.41], $\chi^2_1 = 4.37$, $p < 0.05$. The relationship between visual strength and iconicity ratings was more strongly negative in signed languages.

The model for auditory strength ratings showed that auditory strength was not a significant predictor of iconicity ratings, $\chi^2_1 = 0.03$, n.s. However, there was a significant interaction between auditory strength and modality, $b = 0.21$, 95% CI = [0.10, 0.33], $\chi^2_1 = 7.52$, $p < 0.01$. This revealed that the positive relationship between auditory strength and iconicity ratings was stronger in spoken languages.

The model for haptic strength ratings showed that haptic strength was a significant predictor of iconicity ratings, $b = 0.14$, 95% CI = [0.03, 0.25], $\chi^2_1 = 5.10$, $p < 0.05$. Meanings with greater haptic strength were associated with more iconic signs and words. There was a marginally significant interaction between haptic strength and modality, $b = -0.21$, 95% CI = [-0.42, 0.00], $\chi^2_1 = 3.82$, $p = 0.05$, suggesting that the relationship between haptic strength and iconicity was stronger in signed languages.

The model for gustatory strength ratings showed that gustatory strength was not a significant predictor of iconicity ratings, $\chi^2_1 = 0.24$, n.s., and there was no interaction with modality, $\chi^2_1 = 0.14$, n.s.

Finally, the model for olfactory strength ratings showed that olfactory strength was a significant predictor of iconicity ratings, $b = -0.10$, 95% CI = [-0.19, -0.02], $\chi^2_1 = 3.14$, $p < 0.05$. Across languages, the olfactory strength of meanings was negatively associated with iconicity ratings. There was not an interaction between olfactory strength and modality, $\chi^2_1 = 0.18$, n.s.

These results reveal several interesting patterns across the four languages in the relationship between iconicity and the semantics of signs and words. One notable finding is that iconicity is strongly associated with the concreteness of meanings in the signed languages, but not in the spoken languages. In comparison, while the correlation between iconicity with both sensory experience and imageability is weaker, it is found across the four languages. The relationship between sensory experience and iconicity in English matches previous results using much of the same data (Sidhu and Pexman, 2017; Winter et al.,

2017a), although Winter et al. (2017a) found the opposite – a negative – relationship between iconicity and imageability. In this latter model, Winter et al. included additional factors, including sensory experience rating, which may have accounted for some of the variance explained by imageability in the present model.

A somewhat counterintuitive result was that ratings of visual perceptual strength were negatively correlated with iconicity ratings in both signed languages. Part of the explanation for this may stem from meanings referring to color (e.g., *red*, *blue*, *black*), which are among the meanings with the strongest visual strength. To examine this possibility, we removed color words from the set, and then retested the model of visual strength ratings as a predictor of iconicity ratings. This showed a reduced, but still significant negative effect of visual strength $b = -0.13$, 95% CI = [-0.24, -0.02], $\chi^2_1 = 5.38$, $p < 0.05$). However, the interaction between visual strength and modality was no longer significant, $\chi^2_1 = 0.89$, $p = \text{n.s.}$ Thus, while visual strength was still negatively correlated with iconicity ratings across the languages, after removing color words, this relationship was weaker overall, particularly within the signed languages.

Along with concreteness, haptic strength proved to be the strongest positive predictor of iconicity ratings, both overall, and especially in signed languages. For signed languages, this is an intuitive finding. The haptic sense is largely channeled through manipulative actions of the hands, and therefore, these meanings may afford a high degree of iconicity in signs. The positive correlation between haptic strength and iconicity in English fits with the similar finding by Winter et al. (2017a), which used mostly the same data. Ongoing work suggests that part of the basis for the high iconicity of tactile words may relate to surface texture, and particularly the dimension of roughness versus smoothness (Winter et al., 2017b). However, Spanish appears to contradict this trend common to English and the two signed languages.

As expected, auditory strength was a strong predictor of iconicity in the spoken languages in particular, with an opposing tendency in ASL and BSL. This again replicates Winter et al. (2017a) for English. These results likely reflect the highly compatible format of the vocal-auditory modality of speech for the iconic representation of sound-related meanings (e.g., Perlman and Cain, 2014; Dingemanse et al., 2015).

Across the four languages, the relationship between gustatory and olfactory strength and iconicity was less consistent. For the signed languages and English, it appears to be, if anything, a somewhat negative relationship. Meanings strongly associated with smell and taste tended to have less iconic forms. Spanish, on the other hand, hints at the opposite: a positive relationship between iconicity and meanings related to smell and taste. These preliminary – and tentative – findings with Spanish are unexpected. Smell and taste are distinct from the sensory modalities primarily involved in signed and spoken communication, which directly involve vision, audition and the kinesthetic sense, vis-a-vis the visual and auditory perception of the sights and sounds of bodily movements. And while meanings related to smell and taste are represented by ideophones across languages, they have been counted as less common (Dingemanse, 2012).

In interpreting these different results, it should be considered that all of the ratings for semantic properties were based on the ratings of English glosses judged by English speakers. Thus, the way these ratings characterize the semantics of the translated ASL and BSL signs and the Spanish words is likely to be inaccurate to a degree. Additionally, as a result of this procedure, more English words were covered by the ratings than were the signs and words of the other languages. Consequently, our inferences about English may be more finely tuned than those for the other languages. Conversely, the fewest items were covered for Spanish, leading to wider margins for error in our estimates.

Iconicity and Lexical Classes

In the next set of analyses, we focused on the 220 meanings for which we had iconicity ratings in all four languages. First, we examined how iconicity varied across the vocabularies of the four languages according to broad semantic categories based on the lexical class of the English gloss. **Figure 2** shows iconicity ratings by lexical class – nouns, verbs, adjectives, and adverbs and grammatical words – for each language, displayed as *z*-scores. To test for differences in iconicity between lexical classes, for each language, we constructed a generalized linear model with lexical class as a predictor of iconicity rating.

For ASL, the model showed that nouns were less iconic than verbs, $b = 0.86$, 95% CI = [0.25, 1.48], $t = 2.78$, $p < 0.01$, but more iconic than adjectives, $b = -0.86$, 95% CI = [-1.57, -0.15], $t = -2.37$, $p < 0.05$. Nouns were not significantly higher in iconicity than grammatical words (and adverbs), $b = -0.65$, 95% CI = [-1.48, 0.19], n.s. Similarly, for BSL, nouns were also lower in iconicity than verbs, $b = 0.62$, 95% CI = [0.04, 1.21], $t = 2.10$, $p < 0.05$, but higher than adjectives, $b = -0.78$, 95% CI = [-1.45, -0.10], $t = -2.24$, $p < 0.05$. Again, nouns were not significantly more iconic than grammatical words, $b = -0.51$, 95% CI = [-1.31, 0.28], n.s. For English, nouns were lower in iconicity than both verbs, $b = 0.49$, 95% CI = [0.20, 0.78], $t = 3.27$, $p < 0.01$, and adjectives, $b = 0.58$, 95% CI = [0.24, 0.93], $t = 3.46$, $p < 0.001$, but there was no significant difference between nouns and grammatical words, $b = -0.13$, 95% CI = [-0.53, 0.27],

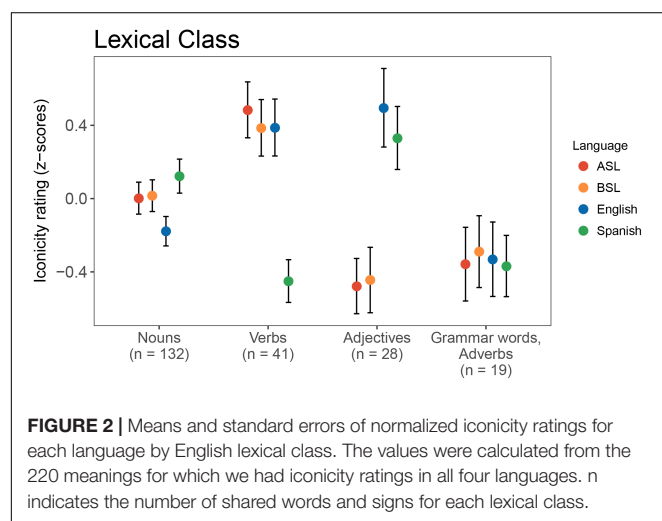
n.s. And finally, the model for Spanish indicated no statistical difference in iconicity between nouns and adjectives, $b = 0.17$, 95% CI = [-0.15, 0.48], n.s., but nouns (and adjectives) were higher in iconicity than verbs, $b = -0.46$, 95% CI = [-0.73, -0.19], $t = -3.24$, $p = 0.001$. Nouns were also significantly higher in iconicity than grammatical words, $b = -0.39$, 95% CI = [-0.77, -0.02], $t = -2.06$, $p < 0.05$.

To determine whether there was an interaction between modality and lexical class, we constructed a linear mixed-effects model of iconicity rating. The model included main effects for (centered) modality and lexical class and a term for their interaction. Random intercepts were included for language and meaning. The results showed a main effect for lexical class, $\chi^2_1 = 14.75$, $p < 0.01$, indicating that overall, nouns were less iconic than verbs, $b = 0.38$, 95% CI = [0.08, 0.68], $t = 2.52$, and more iconic than grammatical words, $b = -0.42$, 95% CI = [-0.83, -0.02], $t = -2.04$, but not more iconic than adjectives, $b = -0.22$, 95% CI = [-0.57, 0.12], $t = -1.26$. There was a highly significant interaction between modality and lexical class, $\chi^2_1 = 43.97$, $p < 0.001$. This interaction reflected that adjectives were relatively higher in iconicity in spoken languages, $b = 1.19$, 95% CI = [0.48, 1.68], $t = 4.86$, and that verbs were higher in signed languages, $b = -0.73$, 95% CI = [-1.14, -0.31], $t = -3.45$. There was no evidence that the iconicity of grammatical words differed between modalities, $b = 0.32$, 95% CI = [-0.25, 0.87], $t = 1.10$.

These analyses point to some interesting differences between signed and spoken languages in how iconicity is spread across broad semantic categories of signs and words. In signed languages, verbs – and thus, presumably, actions – were consistently high in iconicity. This may derive from the natural correspondence between sign and action, as signs are themselves comprised of manual and bodily actions (Armstrong and Wilcox, 2007). Like Perry et al. (2015), we found that English verbs were also high in iconicity, while Spanish verbs were markedly lower. English verbs may be more iconic because they tend to express information about the manner of motion, in contrast to Spanish verbs which do not. Manner of motion might be especially amenable to iconic expression in speech, as, for example, reflected in ideophones (Imai and Kita, 2014).

Notably, in all four languages, nouns – which typically refer to various kinds of things – exhibited an average level of iconicity. Previous work in BSL has suggested that signs for objects, along with actions, are more likely to be iconic (Perniss et al., 2017). Yet, the current results suggest that signs for actions, on the whole, tend to be more iconic than signs for things. At least part of the explanation for this discrepancy may be that there are considerably more nouns in our analyses than other lexical classes. Thus, the nouns may extend to more abstract and complex meanings that are less well suited to iconicity.

As in previous studies (Perry et al., 2015; Winter et al., 2017a), adjectives, which contain many meanings for properties, are rated high in iconicity in both English and Spanish. This contrasts to ASL and BSL, in which adjectives were relatively low in iconicity, at least as compared to nouns and especially to verbs. Similarly, Perniss et al. (2017) found that BSL signs for properties tended not to be iconic. Such findings may be seen to



fall out of line with accounts such as Dingemanse et al. (2015), proposing that meanings related to certain properties – such as ‘size,’ ‘repetition,’ ‘temporal unfolding,’ and ‘intensity’ – may lend themselves to iconicity in both modalities. One possible reason is that the iconicity for the apparently low degree of iconicity in signs for properties is that the iconicity of signed vocabularies is dominated by even more easily representable actions, and to a lesser degree, things. Or it may be that many properties (e.g., colors) do not, in fact, readily lend themselves to iconicity through the manual movements of signs.

Finally, we observed that the miscellaneous category of grammatical words and adverbs tended to be relatively low in iconicity across both the signed and spoken languages. This conclusion is limited by the smaller sample of these meanings, but replicates previous results in English from Perry et al. (2015) and Winter et al. (2017a) with much of the same data. It fits with the prediction of Dingemanse et al. (2015) that meanings like ‘abstract concepts’ and ‘logical operators’ may be hard for both types of languages to represent with iconic forms.

Iconicity and Specific Semantic Categories

Finally, we zoomed in and looked at iconicity across more specific semantic categories for the same 220 meanings. The top panel of

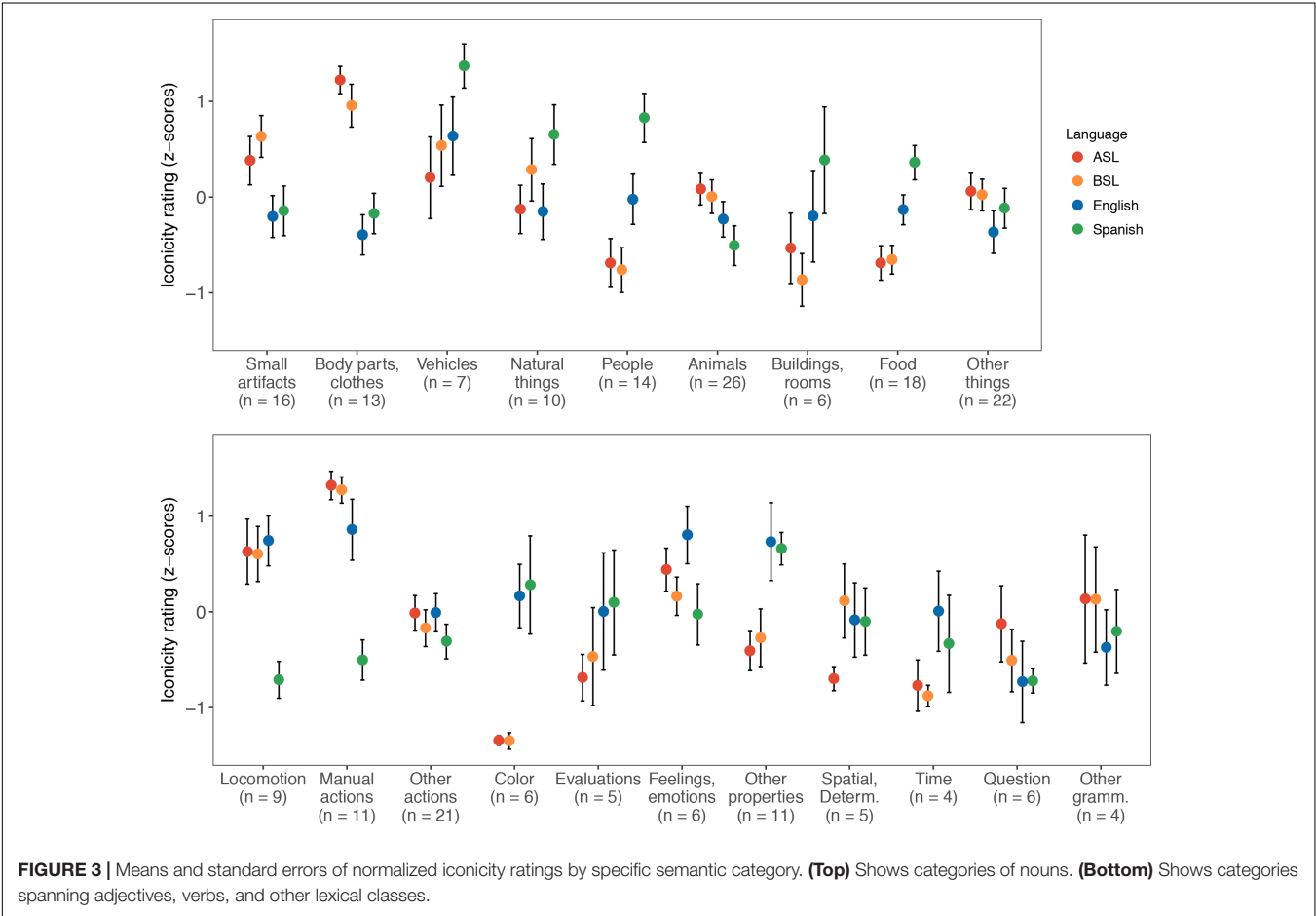
Figure 3 shows the means and standard errors of the z-scored iconicity ratings for semantic categories of nouns, and the bottom panel shows these values for categories derived from adjectives, verbs, and the class of grammatical words and adverbs. Specific examples of words with high, low, and mixed iconicity across signed and spoken languages are presented in **Table 3**.

As shown in the figure, among nouns in the signed languages, small artifacts (i.e., those that can be manipulated by the hands), body parts, and clothes are highest in iconicity, and among verbs, manual actions are especially high. These results demonstrate that in signed languages, iconicity is elevated in meanings related

TABLE 3 | Examples of meanings with high and low iconicity in signed and spoken languages.

Signed iconicity	Spoken iconicity	Meaning
High	High	Baby, elephant, balloon, mouth
High	Low	Book, shoulder, eat, skirt
Low	High	Wet, sick, full, slow, hard
Low	Low	Yellow, doll, apple, and
High	Mixed	Fly, pull, push, break

The row with ‘mixed’ spoken iconicity displays meanings with high iconicity in English, but low iconicity in Spanish.



to the hands and other body parts, supporting the observation of Meir et al. (2013) that the body is ideally suited to represent itself and its parts (also Taub, 2001). On the other end of the scale, iconicity is extremely low in signs for colors. As noted above, the low iconicity of colors contributes to the negative correlation between iconicity and visual strength. Iconicity was also low in signs for time-related meanings and evaluative adjectives, as well as food, buildings and rooms, and terms for different kinds of people (including familial relationships and occupations).

In the spoken languages, over all the noun categories, Spanish words were consistently higher in iconicity than English words, with Spanish nouns for people being especially iconic. Iconicity was highest in nouns for vehicles in both English and Spanish. Outside of nouns, words for other properties were highest in iconicity in both spoken languages. Iconicity in English was also high for feelings and emotions, although this was not the case in Spanish. These results hone previous findings that adjectives, as a broad lexical class, tend to be more iconic in spoken languages, and they fit with cross-linguistic studies showing that ideophones tend to express sensory meanings (Dingemanse, 2012). Of verbs, manual actions and verbs of locomotion were highly iconic in English, but not in Spanish. This pattern may reflect a further refinement of the typological preference of English verbs to express information about manner of motion, which may be more easily rendered into iconic word forms.

Considered together, these findings suggest specific ways in which some semantic categories are more iconic in signed languages, while others are more iconic in spoken languages. Thus, they illustrate the important role of modality in determining how iconicity is distributed across the lexicon of a language. Prominently, the isomorphism between gestures and manual actions appears to motivate a heightened level of iconicity for signs mapping the two (cf. Streeck, 2009). For comparison, however, meanings related to vocal tract actions – that is, those that would afford the spoken parallel of this isomorphism (i.e., onomatopoeia) – are not represented among the 220 meanings we analyzed. Although words for sound-related meanings are generally prevalent in spoken languages, and included in our ratings for English and Spanish, vocabulary to talk about sound is presumably much less common in signed languages. Nevertheless, previous research has shown that, as a domain, sound-related words do tend to be highly iconic (Dingemanse et al., 2016; Winter et al., 2017a).

GENERAL DISCUSSION

Considerable evidence now shows that languages of all sorts, signed and spoken, exhibit iconicity, or resemblance between form and meaning (Perniss et al., 2010; Dingemanse et al., 2015; Perry et al., 2015). From a typological and comparative perspective (e.g., Voeltz and Kilian-Hatz, 2001; Kita and Özyürek, 2003; Dingemanse, 2012; Padden et al., 2013), this raises a host of exciting new questions regarding how iconicity is distributed across the lexicons of different languages. Some of the most basic questions to be answered relate to the modality of the language. Are signed languages really more iconic than spoken

languages? How does the modality of a language influence which lexical forms are iconic and which are not? To investigate these questions, we used previously collected iconicity ratings of signs and words to compare iconicity in the vocabularies of BSL and ASL with those of English and Spanish. Our analyses produced four main sets of findings that serve to characterize how iconicity is spread across the vocabularies of the four languages. These patterns include both interesting similarities between signed and spoken languages, as well as differences between them.

First, we found positive correlations between the iconicity ratings of all four languages, including between English and both ASL and BSL, and between English and Spanish. The one notable exception to this pattern was between Spanish and both of the signed languages – perhaps reflecting the distinctly non-iconic character of Spanish verbs, which tend not to express information about the manner of movement. The relationship between the iconicity ratings of ASL and BSL was especially strong, particular in comparison to that between English and Spanish. This may indicate that the iconicity of signs is, on the whole, more direct and transparent than the iconicity of words – a point to which we return below.

Second, we found that iconicity is distributed over signs and words in systematic ways according to an array of semantic properties. On the whole, signs and words related to the senses – meanings that are more imageable and more connected to sensory experience – are likely to be more iconic. Critically though, concreteness is only associated with more iconicity in signs, not words. Such an asymmetry makes sense, as manual gestures may provide a more concrete semiotic resource for iconicity than do vocalizations. In both types of languages, iconicity is strongest for lexical items with sensory meanings corresponding to the respective language modality – touch in signed languages, and sound in spoken languages.

Third, we found that lexical items for some semantic domains tend to be higher in iconicity than others, and there are characteristic patterns that distinguish between signed and spoken languages. These patterns of iconicity are found at the level of broad semantic categories – for example, actions, things, and properties, as reflected by English glosses as verbs, nouns, and adjectives, respectively. They are also found at the level of more specific semantic categories – manual actions, clothes, emotions, and colors, for example. For the most part, these patterns fit with predictions derived from rationale regarding the semiotic resources of sign versus speech (cf. Dingemanse et al., 2015). For example, in signed languages, signs for actions, and particularly manual actions, are quite high in iconicity, while in spoken languages, words for properties tend to be higher. Critically, this set of analyses was restricted to the 220 meanings with ratings in all four languages, and so the differences between languages cannot be attributed to differential coverage of the ratings.

Finally, one somewhat unexpected set of findings was the relatively low iconicity of nouns and visual words in signed languages, particularly those lacking connection to manual manipulation and the body. While the domain of color was an extreme case of this pattern, it does not provide the complete account. An additional explanation may be that many signs for objects may actually be limited in the level of iconicity possible,

especially in comparison to the iconicity afforded by actions. For example, there may be a certain degree of abstractness involved in using the hands to represent different kinds of things, particularly those that are highly visual. This point is illustrated by the example of the ASL sign for ‘diploma’ (Taub, 2001), which combines the two hands with round handshapes that trace its rolled-up shape. Taub observed that the same iconic resources are modified to represent different kinds of cylinders – water pipes, batons, or a rolled-up poster. Although this scheme makes a productive iconic device, it also demonstrates a baseline of abstractness that derives from mapping the hands to other kinds of objects. This may drive a more moderate level of iconicity for many object meanings, even those with characteristic shapes that can be modeled with the hands.

These four sets of findings point to some interesting new directions for future research into how iconicity is distributed across different kinds of languages. However, it is important to emphasize that our conclusions are preliminary and tentative, and they should be weighed against some notable limitations of our methodology. For one, our study relied opportunistically on samples of rated signs and words that were not originally selected for cross-modal comparison. As a consequence, the ratings that overlapped across languages were somewhat lacking in systematic coverage of the semantic domains that might be of most interest. Additionally, the sample of four languages was not especially well suited to cross-modal comparison. English and Spanish – both Indo-European languages with heavy Latinate influence – are hardly representative of spoken languages. Likewise, ASL and BSL – two widely used, urban signed languages – do not represent the diversity of signed languages (de Vos and Pfau, 2015). Another notable limitation of the study is the disproportionate influence of English on our data. The norms for semantic variables were based on English glosses, rated by English speakers, and similarly, the designations of lexical class were based on English. The iconicity ratings for ASL, and many of those for BSL, were provided by non-signing English speakers. Moreover, the iconicity ratings for Spanish were provided by native speakers of Spanish who were also likely bilingual English speakers, as they were residents of the United States. Given these different factors, it is likely that more diversity in patterns of iconicity would be found by taking a more English-independent approach to a more diverse sample of signed and spoken languages.

Comparing Iconicity in Signed and Spoken Languages

Our findings provide a preliminary, quantified account of how iconicity is spread across the lexicons of signed languages in comparison to spoken languages. As we have sought to demonstrate, the use of iconicity ratings provides researchers with a systematic, standardized method to describe how iconicity is distributed across the vocabulary of a language, which enables direct comparisons between different kinds of languages.

Notably, this approach adopts a strong theoretical premise about the nature of signs and words. The premise is not just that many signs and words are iconic, but as Wescott (1971) observed of what he called “iconism,” pointing to a quotation

from Bronowski (1967, 377): “[T]he only realistic question we can ask about a given form is not ‘Is it iconic?’ but rather ‘How iconic is it?’ A measure of support for this theoretical approach to iconicity is borne out by the richness of the current results. The present work – in addition to several previous studies using iconicity ratings (e.g., Thompson et al., 2012; Perry et al., 2015; Caselli and Pyers, 2017; Occhino et al., 2017) – shows that it is useful to think of iconicity as a “substance” that can shape the forms of signs and words to a greater-or-lesser degree (Dingemanse, 2017b).

In particular, where our study breaks new ground is in its direct side-by-side comparison between signed and spoken languages. We suggest that part of the reason for the previous lack of detailed comparative studies between modalities is the widespread assumption that signed languages are far more iconic than spoken languages (Meier, 2002; Armstrong and Wilcox, 2007). For example, based largely on their intuition, Klima and Bellugi (1979, p. 21) asserted that the “vocabulary of ASL—and, to our knowledge, that of other primary sign languages—is a great deal more iconic than are the morphemes of spoken languages.” This idea figures prominently in many theories of language evolution that argue that the first symbolic forms must have been built from gestures (e.g., Corballis, 2003; Armstrong and Wilcox, 2007; Tomasello, 2008; Arbib, 2012; Fay et al., 2014). These gesture-first theories depend critically on the premise that signs afford much more iconicity than words.

Such an important claim begs for empirical evidence, and indeed, the high correlation we found here between iconicity ratings of ASL and BSL compared to English and Spanish gives it some initial quantitative support. These results suggest that iconic mappings are more consistently realized in the signs of ASL and BSL compared to the words of English and Spanish. To the extent that one can generalize from these four languages, this may indicate that signed languages are iconic in a qualitatively different – and, specifically, a more widely intuitive way – than spoken languages.

However, this intuitiveness may be limited to a significant extent. Previous research has shown that the iconic mappings of signs are not readily obvious to most naïve viewers. For example, experiments have found that non-signers are quite poor at guessing the meanings of unfamiliar signs (Klima and Bellugi, 1979). Of 90 concrete and abstract nouns from ASL, non-signers could not correctly guess the meaning of 81 of them, with the remaining signs only guessed correctly by a small proportion of participants. Guessing was only a little better when constrained to a forced choice with just five alternatives. Similarly, a more recent study of ASL with a much larger set of signs also found that non-signers were very limited in their ability to guess the meaning of the great majority of them (Emmorey and Sevcikova Sehyr, 2018). This lack of transparency is also evidenced in differences in the judgments of iconicity between signers of different languages: signers rate the signs of their native language as more iconic than those of an unfamiliar signed language (Occhino et al., 2017).

Nevertheless, despite this degree of opacity, in the experiments by Klima and Bellugi (1979), when participants were provided with the meaning of the sign, they were often very consistent in explaining the specific correspondence involved. In line

with this, we found with our multiple sets of iconicity ratings for BSL that the ratings provided by non-signers were highly correlated with those provided by signers (see Materials and Methods), as was the case for ASL (Sevcikova Sehyr et al., 2017) – all with coefficients of $r = 0.82$ or higher. In comparison, Perry et al. (2015) found in two experiments with English a correlation of $r = 0.62$ between the ratings, and in two of experiments with Spanish a correlation of $r = 0.41$. While these experiments were each slightly modified in their procedure, they were alike in using only proficient speakers of the respective languages. This greater consistency in the iconicity ratings of signs may reflect the quality – perhaps reflected in a measure of concreteness – that gives traction to accounts of signed languages that postulate clearly identifiable mappings between distinct formal parameters (e.g., handshape, movement, location) and particular aspects of meaning (e.g., shape, motion, position). Such a semiotic framework has given rise to useful theoretical constructs such as structure mapping (Emmorey, 2016) and the double mapping constraint on metaphoric signs (Taub, 2001; Meir, 2010).

Yet, while iconic mappings may be more concrete and structured in signs, spoken languages do still feature their share of transparent mappings. These are found, for example, in correspondences between vowel position and size, between the vocal tract and related anatomical meanings, and between reduplication and iterative action, among others. Clearly structured mappings are especially apparent in the case of onomatopoeia, where there is potential for more isomorphic correspondence between the sound segments of a word and the properties of the sound to which it refers (e.g., Rhodes, 1994).

In addition to highly structured mappings, as our methodology highlights, signs and words may also reflect more abstract and impressionistic correspondences between form and meaning – a vaguer sense that a form looks or sounds like what it means. Thus, it may be that the iconic mappings of signed vocabularies are, on the whole, more concrete and structured, while those of spoken vocabularies are more abstract and impressionistic. Future research – using more nuanced semantic analysis to compare the iconicity of more strategically constructed samples of languages and coverage of vocabulary items – should examine this hypothesis, along with the general

claim that signed languages exhibit higher overall levels of iconicity.

CONCLUSION

Iconicity is now widely documented across the diverse languages of the world, signed and spoken. In both modalities, it is implicated in how people process, produce, and learn to use language, and in the evolutionary processes by which languages are created and change over time. Even if signed languages prove to be “more” iconic than spoken languages, it is becoming clear that this sort of broad generalization is no longer sufficient. Rather, we should aim to describe and compare the detailed, characteristic ways that iconicity is distributed across both kinds of languages.

Although the current study has focused on signs and words, the influence of iconicity extends far beyond the level of the lexicon. Iconicity is also pervasive in the grammars of signed and spoken languages (e.g., Givón, 1985; Liddell, 2002; Aronoff et al., 2005), and in the prosodic inflections (e.g., Shintel et al., 2006; Perlman et al., 2015; Tzeng et al., 2017) and the spontaneous bodily, oral, and vocal gestures that are deeply intertwined with signing and speaking (e.g., McNeill, 1992; Emmorey, 1999; Sandler, 2009; Kendon, 2014; Blackwell et al., 2015; Clark, 2016). Thus, a complete theory of iconicity must seek to explain how the modality of a language figures into the complex interplay between iconicity and the lexicon, as well as all of the other various levels and forms of expression that people use to communicate meaning. Only through such a comprehensive theory of iconicity will we be able to fully understand the nature of human language.

AUTHOR CONTRIBUTIONS

MP primarily conducted the analyses and drafted the manuscript. HL contributed to analyses and helped to revise the manuscript. BT contributed to analysis and helped to draft and revise the manuscript. RLT advised to analyses and helped to revise the manuscript.

REFERENCES

- Aitchison, J. (2007). *The Articulate Mammal: An Introduction to Psycholinguistics*. Abingdon: Routledge.
- Arbib, M. A. (2012). *How the Brain got Language: The Mirror System Hypothesis*. Oxford: Oxford University Press. doi: 10.1093/acprof:osobl/9780199896684.001.0001
- Armstrong, D. F., Stokoe, W. C., and Wilcox, S. E. (1995). *Gesture and the Nature of Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620911
- Armstrong, D. F., and Wilcox, S. E. (2007). *The Gestural Origin of Language*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780195163483.001.0001
- Aronoff, M., Meir, I., and Sandler, W. (2005). The paradox of sign language morphology. *Language* 81, 301–344. doi: 10.1353/lan.2005.0043
- Beavers, J., Levin, B., and Wei Tham, S. (2010). The typology of motion expressions revisited. *J. Linguist.* 46, 331–377. doi: 10.1017/S0022226709990272
- Blackwell, N. L., Perlman, M., and Fox Tree, J. E. (2015). Quotation as a multimodal construction. *J. Pragmat.* 81, 1–7. doi: 10.1016/j.pragma.2015.03.004
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (2016). Sound–meaning association biases evidenced across thousands of languages. *Proc. Natl. Acad. Sci. U.S.A.* 113, 10818–10823. doi: 10.1073/pnas.1605782113
- Bronowski, J. (1967). *Human and Animal Languages. To Honor Roman Jakobson*, 1.374–94. The Hague: Mouton.
- Brysbaert, M., Warriner, A. M., and Kuperman, V. (2014). Concreteness ratings for 40 thousands generally known English word lemmas. *Behav. Res. Methods* 46, 904–911. doi: 10.3758/s13428-013-0403-5
- Brysbaert, M., and Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behav. Res. Methods* 44, 991–997. doi: 10.3758/s13428-012-0190-4
- Cartmill, E. A., Beilock, S., and Goldin-Meadow, S. (2012). A word in the hand: action, gesture and mental representation in humans and non-human

- primates. *Philos. Trans. R. Soc. B* 367, 129–143. doi: 10.1098/rstb.2011.0162
- Caselli, N., Sevcikova, Z., Cohen-Goldberg, A., and Emmorey, K. (2017). ASL-Lex: a lexical database for ASL. *Behav. Res. Methods* 49, 784–801. doi: 10.3758/s13428-016-0742-0
- Caselli, N. K., and Pyers, J. E. (2017). The road to language learning is not entirely iconic: iconicity, neighborhood density, and frequency facilitate acquisition of sign language. *Psychol. Sci.* 28, 979–987. doi: 10.1177/0956797617700498
- Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123, 324–347. doi: 10.1037/rev0000026
- Corballis, M. C. (2003). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.
- Cortese, M. J., and Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behav. Res. Methods Instrum. Comput.* 36, 384–387. doi: 10.3758/BF03195585
- de Saussure, F. (1959). *Course in General Linguistics*. New York, NY: Philosophical Library.
- de Vos, C., and Pfau, R. (2015). Sign language typology: the contribution of rural sign languages. *Annu. Rev. Linguist.* 1, 265–288. doi: 10.1146/annurev-linguist-030514-124958
- Diffloth, G. (1972). Notes on expressive meaning. *Paper Presented at the 8th Regional Meeting of the Chicago Linguistic Society*. Chicago, IL.
- Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Lang. Linguist. Compass* 6, 654–672. doi: 10.1002/lnc3.361
- Dingemanse, M. (2014). Making new ideophones in Siwu: creative depiction in conversation. *Pragmat. Soc.* 5, 384–405. doi: 10.1075/ps.5.3.04din
- Dingemanse, M. (2017a). Expressiveness and system integration: on the typology of ideophones, with special reference to Siwu. *STUF Lang. Typol. Univers.* 70, 363–384. doi: 10.1515/stuf-2017-0018
- Dingemanse, M. (2017b). Operationalising iconicity. *Paper Presented at 11th International Symposium on Iconicity in Language & Literature, April 8, Brighton*.
- Dingemanse, M., and Akita, K. (2017). An inverse relation between expressiveness and grammatical integration: on the morphosyntactic typology of ideophones, with special reference to Japanese 1. *J. Linguist.* 53, 501–532. doi: 10.1017/S00222671600030X
- Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends Cogn. Sci.* 19, 603–615. doi: 10.1016/j.tics.2015.07.013
- Dingemanse, M., Schuermer, W., Reinisch, E., Tufvesson, S., and Mitterer, H. (2016). What sound symbolism can and cannot do: testing the iconicity of ideophones from five languages. *Language* 92, e117–e133. doi: 10.1353/lan.2016.0034
- Emmorey, K. (1999). “Do signers gesture?” in *Gesture, speech, and sign*, eds L. S. Messing and R. Campbell (Oxford: Oxford University Press), 133–159. doi: 10.1093/acprof:oso/9780198524519.003.0008
- Emmorey, K. (2016). Consequences of the Now-or-Never bottleneck for signed versus spoken languages. *Behav. Brain Sci.* 39:e70 doi: 10.1017/S0140525X1500076X
- Emmorey, K., and Sevcikova Sehyr, Z. (2018). Measuring iconicity in sign language: effects of linguistic knowledge and transparency on perceived iconicity. *Paper Presented at the International Society for Gesture Studies Conference, July 8, Cape Town*.
- Fay, N., Lister, C. J., Ellison, T. M., and Goldin-Meadow, S. (2014). Creating a communication system from scratch: gesture beats vocalization hands down. *Front. Psychol.* 5:354 doi: 10.3389/fpsyg.2014.00354
- Gasser, M. (2004). “The origins of arbitrariness in language,” in *Proceedings of the 26th Annual Conference of the Cognitive Science Society*, (Mahwah, NJ: Lawrence Erlbaum), 434–439.
- Givón, T. (1985). “Iconicity, isomorphism, and non-arbitrary coding in syntax,” in *Iconicity in Syntax*, ed. J. Haiman (Amsterdam: John Benjamins Publishing).
- Givón, T. (2001). *Syntax: A Functional-Typological Introduction*, Vol. 1. Amsterdam: Benjamins. doi: 10.1075/z.syn1
- Haynie, H., Bower, C., and LaPalombara, H. (2014). Sound symbolism in the languages of Australia. *PLoS One* 9:e92852. doi: 10.1371/journal.pone.0092852
- Hockett, C. F. (1960). The origin of speech. *Sci. Am.* 203, 88–97. doi: 10.1038/scientificamerican0960-88
- Hockett, C. F. (1978). In search of Jove’s brow. *Am. Speech* 53, 243–313. doi: 10.2307/455140
- Imai, M., and Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130298. doi: 10.1098/rstb.2013.0298
- Johansson, N., and Zlatev, J. (2013). Motivations for sound symbolism in spatial deixis: a typological study of 101 languages. *Public J. Semiot.* 5, 3–20.
- Johnston, T., and Schembri, A. (1999). On defining lexeme in a signed language. *Sign Lang. Linguist.* 2, 115–185. doi: 10.1075/sll.2.2.03joh
- Juhász, B. J., and Yap, M. J. (2013). Sensory experience ratings for over 5,000 mono- and disyllabic words. *Behav. Res. Methods* 45, 160–168. doi: 10.3758/s13428-012-0242-9
- Kantantzis, K., Imai, M., and Kita, S. (2011). Japanese sound-symbolism facilitates word learning in English-speaking children. *Cogn. Sci.* 35, 575–586. doi: 10.1111/j.1551-6709.2010.01169.x
- Kendon, A. (2014). Semiotic diversity in utterance production and the concept of ‘language’. *Philos. Trans. R. Soc. B* 369:20130293. doi: 10.1098/rstb.2013.0293
- Kita, S., and Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: evidence for an interface representation of spatial thinking and speaking. *J. Mem. Lang.* 48, 16–32. doi: 10.1016/S0749-596X(02)00505-3
- Klima, E., and Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.
- Kwon, N. (2015). *The Natural Motivation of Sound Symbolism*. Ph.D. thesis, Brisbane, University of Queensland.
- Kwon, N. (2017). Empirically observed iconicity levels of English phonaestemes. *Public J. Semiot.* 7, 73–93.
- Langacker, R. W. (2008). *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780195331967.001.0001
- Lemaitre, G., Houix, O., Voisin, F., Misdariis, N., and Susini, P. (2016). Vocal imitations of non-vocal sounds. *PLoS One* 11:e0168167. doi: 10.1371/journal.pone.0168167
- Lepic, R., Børstell, C., Belsitzman, G., and Sandler, W. (2016). Taking meaning in hand: iconic motivations in two-handed signs. *Sign Lang. Linguist.* 19, 37–81. doi: 10.1075/sll.19.1.02lep
- Liddell, S. K. (2002). “Modality effects and conflicting agendas,” in *The Study of Signed Languages: Essays in Honor of William C. Stokoe*, eds D. F. Armstrong, M. A. Karchmer, and J. V. Van Cleve (Washington, DC: Gallaudet University Press), 53–81.
- Lynott, D., and Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behav. Res. Methods* 41, 558–564. doi: 10.3758/BRM.41.2.558
- Lynott, D., and Connell, L. (2013). Modality exclusivity norms for 400 nouns: the relationship between perceptual experience and surface word form. *Behav. Res. Methods* 45, 516–526. doi: 10.3758/s13428-012-0267-0
- Massaro, D. W., and Perlman, M. (2017). Quantifying iconicity’s contribution during language acquisition: implications for vocabulary learning. *Front. Commun.* 2:4. doi: 10.3389/fcomm.2017.00004
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago press.
- Meier, R. P. (2002). “Why different, why the same? Explaining effects and non-effects of modality upon linguistic structure in sign and speech,” in *Modality and Structure in Signed and Spoken Languages*, eds R. R. Meier, K. Cormier, and D. Quinto Pozos (Cambridge: Cambridge University Press), 1–25.
- Meir, I. (2010). Iconicity and metaphor: constraints on metaphorical extension of iconic forms. *Language* 86, 865–896. doi: 10.1353/lan.2010.0044
- Meir, I., Padden, C., Aronoff, M., and Sandler, W. (2013). Competing iconicities in the structure of languages. *Cogn. Linguist.* 24, 309–343. doi: 10.1515/cog-2013-0010
- Meir, I. and Tkachman, O. (2018). “Iconicity,” in *Oxford Research Encyclopedia of Linguistics*, ed. M. Aronoff (Oxford: Oxford University Press). doi: 10.1093/acrefore/9780199384655.013.343
- Newmeyer, F. J. (1992). Iconicity and generative grammar. *Language* 68, 756–796. doi: 10.1353/lan.1992.0047
- Nuckolls, J. B. (1999). The case for sound symbolism. *Annu. Rev. Anthropol.* 28, 255–282. doi: 10.1146/annurev.anthro.28.1.225
- Occhino, C., Anible, B., Wilkinson, E., and Morford, J. P. (2017). Iconicity is in the eye of the beholder. *Gesture* 16, 100–126. doi: 10.1075/gest.16.1.04occ

- Ohala, J. J. ed. (1994). "The frequency code underlies the sound-symbolic use of voice pitch," in *Sound Symbolism*, (Cambridge: Cambridge University Press), 325–347.
- Östling, R., Börstell, C., and Courtaux, S. (2018). Visual iconicity across sign languages: large-scale automated visual analysis of iconic articulators and locations. *Front. Psychol.* 9:725. doi: 10.3389/fpsyg.2018.00725
- Padden, C. A., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S., and Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture* 13, 287–308. doi: 10.1075/gest.13.3.03pad
- Perlman, M., and Cain, A. A. (2014). Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture* 14, 320–350. doi: 10.1075/gest.14.3.03per
- Perlman, M., Clark, N., and Johansson Falck, M. (2015). Iconic prosody in story reading. *Cogn. Sci.* 39, 1348–1368. doi: 10.1111/cogs.12190
- Perniss, P., Lu, J. C., Morgan, G., and Vigliocco, G. (2017). Mapping language to the world: the role of iconicity in the sign language input. *Dev. Sci.* 21:e12551. doi: 10.1111/desc.12551
- Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol.* 1:227. doi: 10.3389/fpsyg.2010.00227
- Perniss, P., and Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philos. Trans. Actions R. Soc. B Biol. Sci.* 369:20130300. doi: 10.1098/rstb.2013.0300
- Perry, L. K., Perlman, M., and Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS One* 10:e0137147. doi: 10.1371/journal.pone.0137147
- Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., and Lupyan, G. (2017). Iconicity in the speech of children and adults. *Dev. Sci.* 21:e12572. doi: 10.1111/desc.12572
- Pietrandrea, P. (2002). Iconicity and arbitrariness in Italian sign language. *Sign Lang. Stud.* 2, 296–321. doi: 10.1353/sls.2002.0012
- Pinker, S. (1999). *Words and Rules: the Ingredients of Language*. London: Weidenfeld & Nicolson.
- Pinker, S., and Bloom, P. (1990). Natural language and natural selection. *Behav. Brain Sci.* 13, 707–727. doi: 10.1017/S0140525X00081061
- Pitcher, B. J., Mesoudi, A., and McElligott, A. G. (2013). Sex-biased sound symbolism in English-language first names. *PLoS One* 8:e64825. doi: 10.1371/journal.pone.0064825
- Rhodes, R. (1994). "Aural images," in *Sound Symbolism*, eds L. Hinton, J. Nichols, and J. Ohala (Cambridge: Cambridge University Press), 276–292.
- Sandler, W. (2009). Symbiotic symbolization by hand and mouth in sign language. *Semiotica* 174, 241–275. doi: 10.1515/semi.2009.035
- Sevcikova Sehyr, Z., Fisher, D., and Emmorey, K. (2017). "The influence of sign language experience on iconicity judgments," in *Poster Presented at Psychonomic Society's 58th Annual Meeting*, Vancouver, BC.
- Shintel, H., Nusbaum, H. C., and Okrent, A. (2006). Analog acoustic expression in speech communication. *J. Mem. Lang.* 55, 167–177. doi: 10.1016/j.jml.2006.03.002
- Sidhu D., and Pexman, P. (2017). Lonely sensational icons: semantic neighborhood density, sensory experience and iconicity. *Lang. Cogn. Neurosci.* 33, 25–31. doi: 10.1080/23273798.2017.1358379
- Stokoe, W. C. (1965). *A Dictionary of American Sign Language*. Washington, DC: Gallaudet College Press.
- Streeck, J. (2009). *Gestures: The Manu-Facture of Meaning*. Amsterdam: John Benjamins Publishing Company. doi: 10.1075/gs.2
- Strik Lievers, F., and Winter, B. (2018). Sensory language across lexical categories. *Lingua* 204, 45–61. doi: 10.1016/j.lingua.2017.11.002
- Talmy, L. (2000). *Toward a Cognitive Semantics*, Vol. 2. Cambridge, MA: MIT press.
- Tanz, C. (1971). Sound symbolism in words relating to proximity and distance. *Lang. Speech* 14, 266–276. doi: 10.1177/002383097101400307
- Taub, S. F. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511509629
- Taylor, K. T. (2007). *KA-BOOM! A Dictionary of Comic Book Words, Symbols & Onomatopoeia*. Morrisville, NC: Lulu.com.
- Thompson, R. L., Vinson, D. P., Woll, B., Vigliocco, G. (2012). The road to language learning is iconic evidence from British Sign Language. *Psychol. Sci.* 23, 1443–1448. doi: 10.1177/0956797612459763
- Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.
- Tzeng, C. Y., Duan, J., Namy, L. L., and Nygaard, L. C. (2017). Prosody in speech as a source of information. *Lang. Cogn. Neurosci.* 33, 512–526. doi: 10.1080/23273798.2017.1391400
- Ullan, R. (1978). Size-sound symbolism. *Univ. Hum. Lang.* 2, 525–568.
- Urban, M. (2011). Conventional sound symbolism in terms for organs of speech: a cross-linguistic study. *Folia Linguist.* 45, 199–213. doi: 10.1515/flin.2011.007
- Vigliocco, G., Perniss, P., and Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philos. Trans. R. Soc. B* 369:20130292. doi: 10.1098/rstb.2013.0292
- Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., and Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behav. Res. Methods* 40, 1079–1087. doi: 10.3758/BRM.40.4.1079
- Voeltz, F. E., and Kilian-Hatz, C. (eds). (2001). *Ideophones*, Vol. 44. Amsterdam: John Benjamins Publishing. doi: 10.1075/tsl.44
- Wescott, R. W. (1971). Linguistic iconism. *Language* 47, 416–428. doi: 10.2307/412089
- Whitney, W. D. (1874). Phusei or Thesei—natural or conventional? *Trans. Am. Philol. Assoc.* 6, 95–116.
- Wichmann, S., Holman, E. W., and Brown, C. H. (2010). Sound symbolism in basic vocabulary. *Entropy* 12, 844–858. doi: 10.3390/e12040844
- Wilcox, S. (2004). Cognitive iconicity: conceptual spaces, meaning, and gesture in signed language. *Cogn. Linguist.* 15, 119–148. doi: 10.1515/cogl.2004.005
- Winter, B. (2016). *The Sensory Structure of the English Lexicon*. Ph.D. thesis, University of California, Merced.
- Winter, B., Perlman, M., Perry, L. K., and Lupyan, G. (2017a). Which words are most iconic? Iconicity in English sensory words. *Interact. Stud.* 18, 433–454. doi: 10.1098/rstb.2017.0137
- Winter, B., Sósokuthy, M., and Perlman, M. (2017b). "R' is for rough: iconicity in English and Hungarian surface descriptors," in *Poster Presented at the Iconicity Focus Group Workshop*, Nijmegen.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Perlman, Little, Thompson and Thompson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Creating Images With the Stroke of a Hand: Depiction of Size and Shape in Sign Language

Jenny C. Lu^{1*} and Susan Goldin-Meadow²

¹ Department of Psychology, The University of Chicago, Chicago, IL, United States, ² Departments of Psychology and Comparative Human Development, The University of Chicago, Chicago, IL, United States

OPEN ACCESS

Edited by:

Wendy Sandler,
University of Haifa, Israel

Reviewed by:

Naomi K. Caselli,
Boston University, United States
Rose Stamp,
University of Haifa, Israel

*Correspondence:

Jenny C. Lu
jennylu@uchicago.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 01 January 2018

Accepted: 03 July 2018

Published: 31 July 2018

Citation:

Lu JC and Goldin-Meadow S (2018)
Creating Images With the Stroke of a
Hand: Depiction of Size and Shape
in Sign Language.
Front. Psychol. 9:1276.
doi: 10.3389/fpsyg.2018.01276

In everyday communication, not only do speakers describe, but they also depict. When depicting, speakers take on the role of other people and quote their speech or imitate their actions. In previous work, we developed a paradigm to elicit depictions in speakers. Here we apply this paradigm to signers to explore depiction in the manual modality, with a focus on depiction of the size and shape of objects. We asked signers to describe two objects that could easily be characterized using lexical signs (Descriptive Elicitation), and objects that were more difficult to distinguish using lexical signs, thus encouraging the signers to depict (Depictive Elicitation). We found that signers used two types of depicting constructions (DCs), conventional DCs and embellished DCs. Both conventional and embellished DCs make use of categorical handshapes to identify objects. But embellished DCs also capture imagistic aspects of the objects, either by adding a tracing movement to gradually depict the contours of the object, or by adding a second handshape to depict the configuration of the object. Embellished DCs were more frequent in the Depictive Elicitation context than in the Descriptive Elicitation context; lexical signs showed the reverse pattern; and conventional DCs were equally like in the two contexts. In addition, signers produced iconic mouth movements, which are temporally and semantically integrated with the signs they accompany and depict the size and shape of objects, more often with embellished DCs than with either lexical signs or conventional DCs. Embellished DCs share a number of properties with embedded depictions, constructed action, and constructed dialog in signed and spoken languages. We discuss linguistic constraints on these gradient depictions, focusing on how handshape constrains the type of depictions that can be formed, and the function of depiction in everyday discourse.

Keywords: depiction, depicting constructions, iconic mouth movements, gesture, iconicity

INTRODUCTION

In everyday communication, not only do people use words to convey their thoughts and actions, but they also often iconically demonstrate what they are thinking or seeing. For example, consider two accounts of a bicycle accident:

- (1) The bike crashed into the woman.
- (2) The bike hit the woman, BAAAM [iconic gesture in which one hand collides with the other].

In (1) the speaker describes the event using only conventional lexical items, and conveys the fact that the collision was violent with the word “crash.” In (2) the speaker describes the event with a less evocative lexical item (“hit”) but adds information about the severity of the collision with vocal onomatopoeia and the vowel elongated, *BAAM*, accompanied by an imagistic gesture depicting the crash. Instead of simply describing the event as in (1), the speaker in (2) combines two modes of representations—conventional signs and spontaneous depictions. The second utterance may do a better job of evoking a sensory image of the event, allowing one to imagine just how bad the collision was.

Speakers often combine different communicative devices—words, gestures, enactments—to produce multi-modal “composite utterances” (Enfield, 2009). Depiction is one of these communicative devices often used along with conventional linguistic forms. When depicting, speakers can vocally represent an object or event in an iconic and meaningful way (Clark and Gerrig, 1990). The goal in depiction is to set up a physical scene that is analogous to the real-world scene, and to invite the listener to imagine the sensory or visual experience (Clark and Gerrig, 1990; Clark, 2016). The forms used in a depiction are often unconventional, and map onto meaning gradiently rather than categorically (Shintel et al., 2006). Users of sign language also make use of depiction (Liddell, 2003; Streeck, 2008), innovating visual forms that map gradiently onto meaning (Okrent, 2002; Emmorey and Herzig, 2003). Here we focus on how depiction is used by signers in situations designed to be difficult to describe.

Depiction in Spoken Languages

Depiction in speakers can have a significant semantic, and possibly even grammatical role, in a sentence (Ferrara and Johnston, 2014; Hodge and Johnston, 2014). For example, a depiction can function as a noun or verb phrase, embedded within a larger phrase (Clark, 2016). In example (3), the speaker was discussing a piece played by Bela Bartok. He starts his sentence by saying, “he does not play,” and then depicts a musical passage on the piano using a style not used by Bartok—this passage takes on the role of a noun phrase in the speaker’s sentence. The speaker then contrasts this depiction with another depiction, a musical rendition of how Bartok *did* play the passage, which also takes on the role of a noun phrase.

- (3) He does not play (*demonstrates four measures on the piano while singing*) but rather he plays (*demonstrates the same four measures while singing, but differently*)—and he does it better than I do. Clark (2016)

These musical depictions function as parts of the speaker’s sentence, and the speaker uses them contrastively to highlight how the piece was actually played. Note that the speaker’s sentence is actually incomplete without the two performative chunks. However, even though these depictions are functioning as part of the speaker’s sentence, they are not conventional linguistic forms. In the *Bartok* example (3), the depictive forms were created on the spot by the speaker, but can be immediately understood by others through the iconic mapping between the forms and the events they represent.

Speakers also use depiction to demonstrate the speech, affect, and emotions of another person. In the following example (4), Matt talks about a customer, Beth.

- (4) She says ‘well, I’d like to buy an ant’ Clark and Gerrig (1990)

Matt is not referring to himself when he says “I,” but is instead taking on the role of Beth, who is talking to a store clerk. As such, he may also be raising the pitch of his voice and gesturing as Beth might gesture. Examples of this sort are referred to as role shift, direct quotation, direct speech, or constructed dialog.

Depictions can contain a mixture of categorical and gradient forms. Mimetic words, called ideophones, found in languages like Japanese or Siwu (Kita, 1997; Dingemanse and Akita, 2016), are good examples. These spoken devices are iconic, sensory words that contain properties amenable to gradient modification (e.g., reduplication or vowel lengthening). A Japanese speaker can reduplicate the ideophone *zorot(-te)*, which means ‘one after another in line,’ to create *zorozorot-te*, which iconically expresses the intensity of incoming waves. This kind of expressive foregrounding is not limited to mimetic linguistic forms. Non-mimetic words can be modified analogically by, for example, elongating the vowel “o” in the word *long* in a meaningful way: “It was a *loooooong* time” (Okrent, 2002; Shintel et al., 2006). The meaning of the categorical form is preserved, while the analog acoustic overlay contributes additional rich meaning (Kästner and Newen, 2017).

Depiction in Sign Languages

Depiction is not only prevalent in everyday spoken language, but is also common in signed language. Signers often use constructed dialog or constructed action, where the signer takes on the role of another person or produces an action of another person or object (Metzger, 1995; Quinto-Pozos, 2007; Cormier et al., 2015). In Quinto-Pozos (2007) study eliciting utterances with constructed action, a signer takes on the role of a seal and enacts its behavior by using his body to represent the animal’s body. The hands form 5 (✋) handshapes and are placed on the side of the body to represent the flippers. Then, the head and torso sway forward and back, and the mouth opens and closes as if representing the mouth of seal, which appears to be pure enactment.

In a different example, a signer first identifies an agent by signing WOMAN, and then describes the agent’s goal—MONEY HOW-MUCH TOTAL (5); the woman wanted to know the total cost (Cormier et al., 2012). The signer then enacts the woman’s actions on a calculator by using her body to represent the woman’s body and her hands to represent the woman’s hands on the calculator. The handling handshape representing the agent’s actions on the calculator is considered to be a depicting construction (DC). As in the *Bela Bartok* example above, the depiction completes the sentence and makes it comprehensible.

- (5) WOMAN MONEY HOW-MUCH TOTAL (*enacts using the calculator with handling DC*) Cormier et al. (2012)

Signers also use DCs, also known as classifier constructions. DCs are comparable to Japanese mimetics in the sense that they

are gradient and can represent physical properties of events. In these predicates, handshape is used to denote the category of the object being described; for example, an index finger (☞) handshape represents a long, thin object, or a bent-V (☞) handshape represents an animal (Supalla, 1982, 1986). These forms are conventional, although they can take on mimetic properties (as do ideophones in speech); for example, the handshape can be iconically and topographically moved in sign space to portray the location and motion of the object. In these constructions, handshape is categorical, with clear and predictable mappings onto semantic categories, but movement is arguably gradient (Emmorey and Herzig, 2003; Liddell, 2003; although see Supalla, 1982, for a different view). This gradient property of DCs makes these forms highly productive; signers can combine multiple components and manipulate them gradiently in space in a seemingly infinite number of ways.

The sign language literature typically characterizes a linguistic form as *either* categorical *or* gradient. But it can, in fact, be both (Emmorey and Herzig, 2003). Recent research on Taiwanese Sign Language (e.g., Duncan, 2005) has shown that signers gradiently modify their categorical handshapes, often using these gradient devices to convey the same type of information that speakers incorporate into their co-speech gestures. Signers used an animal classifier handshape (thumb-and-pinky handshape) to represent the cat in a story, and gradiently modified the handshape to represent the cat's ever-changing body form as it moved up the drainpipe. Indeed the signers used the same modifications to capture the cat's movements that hearing speakers use in their co-speech gestures describing the same scene (Duncan, 2005).

The Current Study

The goal of our paper is to characterize depiction in signers with an eye toward similar phenomena in spoken languages. In previous work, we developed a paradigm for eliciting depiction in speakers (Lu and Goldin-Meadow, 2017). Here we apply this paradigm to signers to explore depiction in the manual modality. In a within-subjects design, we asked speakers to describe two objects that could easily be characterized using lexical words in English (Descriptive Elicitation), and two objects that were more difficult to distinguish using English lexical words, thus encouraging speakers to depict (Depictive Elicitation). When speakers struggle to access lexical words, they gesture at relatively high rates (Chawla and Krauss, 1994; Hostetter et al., 2007; Sevcikova Sehyr et al., 2018), and we found that speakers did indeed use more iconic gestures in the Depictive Elicitation context than in the Descriptive Elicitation context (Lu and Goldin-Meadow, 2017).

Here we ask how signers behave in Depictive Elicitation contexts, and focus on the depiction of the size and shape of objects, where handshape and movement contribute to creating meaning about both properties. Depiction of size and shape is a relatively underexplored area compared to depiction of action, handling, or viewpoint (Metzger, 1995; Quinto-Pozos, 2007; Cormier et al., 2015). Previous literature has characterized handshape and movement within DCs as categorical (e.g., Supalla, 1986); we extend this research by exploring other kinds of DCs that have imagistic and gradient qualities.

We developed a paradigm that systematically elicits depictive devices. Signers described to a camera pairs of objects that belonged to the same category and were of the same color but differed in shape (e.g., a yellow vase of shape 1 vs. yellow vase of shape 2). The lexical signs YELLOW and VASE do not distinguish between the two vases, and there are no lexical items in ASL that correspond to the difference in shape between the two yellow vases. As a result, signers may feel the need to depict. We explore the depictive strategies signers use when lexical signs alone are not likely to suffice in communicating the full message.

MATERIALS AND METHODS

Participants

Nineteen deaf participants, fluent in American Sign Language (ASL), were recruited at a local Deaf event or through email advertisement. Four participants were excluded from the data analyses because they did not understand the instructions (e.g., they created elaborate stories unrelated to the task, or did not use any lexical signs in their descriptions; $n = 3$) or because they had proficiency in another sign language prior to learning ASL ($n = 1$). Data from the remaining 15 participants were coded by a deaf and a hearing coder, both of whom were fluent in ASL. The mean age of first exposure to ASL is 0.43 years ($SD = 0.82$, range: 0–3 years), and 10 out of 15 participants were native signers born to deaf parents. Thirteen participants gave ASL as their preferred language; the remaining two did not respond to the question. Signers were paid \$50 for their participation and travel.

The stimuli were 24 pairs of objects presented on a computer screen, 12 in the Depictive Elicitation context (pairs of objects that are difficult to distinguish using lexical signs, e.g., a yellow vase of shape 1 vs. a yellow vase of shape 2; **Figure 1**; Supplementary Material), and 12 in the Descriptive Elicitation context (pairs of objects that could be identified by different lexical signs, e.g., pot vs. bowl)¹. The presentation of stimuli was programmed on Psyscope X B77 (Cohen et al., 1993). The objects in both the Descriptive and Depictive Elicitation contexts belonged to five different shape categories: (1) long and thin objects (e.g., a stick), (2) small and discrete objects (e.g., pills), (3) cylindrical objects (e.g., a vase), (4) round objects (e.g., a rock), and (5) objects with a combination of shapes, (e.g., a mushroom, which was a combination of a thin stem and a round cap). The contexts were designed to elicit the following handshapes—Claw 5 (☞), C (☞), F (☞) handshapes (see Supplementary Tables S1, S2). The position of the objects appearing on the left or right side of the screen was counterbalanced, and the order of presentation of stimuli was random.

Procedure

Prior to the study, participants were interviewed online about their language and education background, and also filled out consent forms and a questionnaire on their language background.

¹Participants also saw 12 pairs of objects that were from the same category but of a different color (e.g., a white ornament vs. a green ornament) and thus could be identified by different color signs. The data from this 'Descriptive (color)' condition were not analyzed here.

Deaf participants were asked in ASL by a native deaf signer to describe what they saw on the screen; they were told that this was a communication task and that another participant would later watch the video of their responses and be asked to identify which of the two objects the response referred to. Once they completed their responses, the experimenter debriefed the participants on the goal of the study.

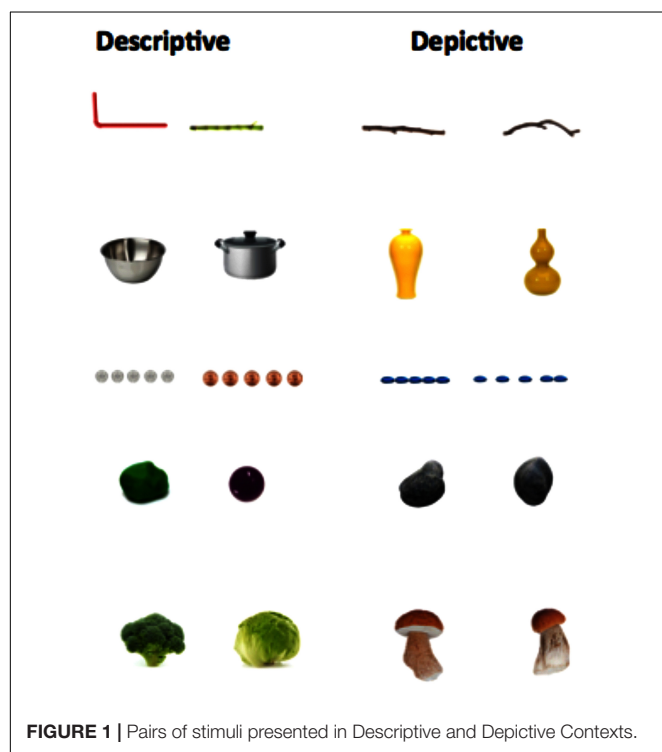


FIGURE 1 | Pairs of stimuli presented in Descriptive and Depictive Contexts.

TABLE 1 | Total number of Conventional and Embellished DCs produced by each participant in the Descriptive and Depictive Elicitation Conditions.

Participant	Descriptive condition		Depictive condition	
	Conventional DCs	Embellished DCs	Conventional DCs	Embellished DCs
1	12	14	10	19
2	1	7	8	16
3	8	21	33	23
4	2	14	6	28
5	12	14	9	42
6	31	23	5	27
7	3	14	3	30
8	15	26	6	41
9	10	33	9	42
10	40	58	5	51
11	0	12	3	40
12	24	56	10	51
13	5	60	5	58
14	37	46	39	73
15	15	87	14	91

Coding

We transcribed all of the lexical items and DCs that each participant produced in the Depictive and Descriptive Elicitation contexts. We included in the analyses core lexical signs (Brentari and Padden, 2001) and fingerspelled words serving as nouns or adjectives, as well as DCs describing perceptual attributes of the objects (see **Figure 2**). In a lexical sign, the handshape, location, and movement are fixed (unless inflected by a regular morphological process).

Depicting constructions are also called “classifiers” (Frishberg, 1975; Kegl and Wilbur, 1976; Supalla, 1986), “polymorphemic signs” (Engberg-Pederson, 1993), “polycomponential signs” (Slobin et al., 2003), and “depicting verbs” (Liddell, 2003). Following Johnston and Schembri (2010) and Cormier et al. (2012), we use the term “depicting constructions,” DCs. Each component of a DC—handshape, location, movement—bears meaning (unlike phonological components of signs), and each is a bound morpheme that can recombine with each other. We excluded the few DCs that were used to describe actions performed on or with the objects (e.g., handling DCs), as well as pointing and numbers signs. We identified two main types of DCs in our data: conventional DCs and embellished DCs.

Conventional DCs

These DCs are also known as “size and shape specifiers” (SASS) or, more recently, “entity DCs” (Cormier et al., 2012; Zwitserlood, 2012). The handshape in these DCs represents the shape of the object (Supalla, 1986). For example, a signer could use an F (✋) handshape to represent a coin, a C (✋) handshape to represent a bottle, or an index finger (✋) handshape to represent a pen. The handshapes in conventional DCs are fixed but, unlike lexical signs, can combine with a variety of movements or locations. However, the handshapes in the conventional DCs produced in our study were combined, for the most part, with a hold movement, or a series of holds, in neutral space. If the signer produced a series of repeated handshapes, without pausing, to indicate a set of items (e.g., pills), this response was coded as a single conventional DC. If the signer described the first three pills in a row, paused, and then described the second two pills (which were spaced closer together than the other three), this response was coded as two separate conventional DCs (see **Figure 2**).

Embellished DCs

Participants also produced DCs that have imagistic components added to a conventional base. Conventional DCs draw from a conventional set of handshapes (Supalla, 1986; Zwitserlood, 2003). Embellished DCs use these same handshapes but embellish them, either by adding a second conventional handshape or by adding movement. These embellished DCs appear to be spontaneously created at the moment to capture particular aspects of the objects to which they refer.

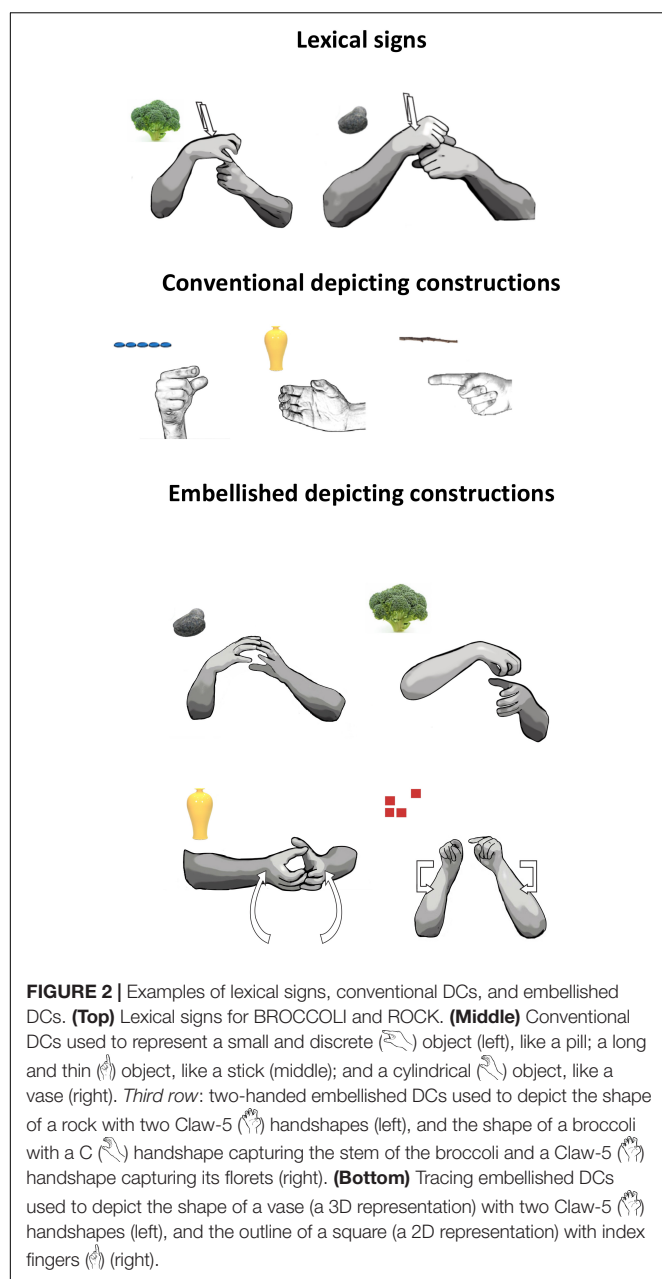
Combining two handshapes

Combining two conventional handshapes allows signers to capture objects with a complex configuration or with multiple parts (see Sevcikova Sehyr et al., 2018). For example, a signer can use a C (✋) handshape on the non-dominant hand to represent

TABLE 2 | Number of conventional and embellished DCs produced by each participant for five representative stimuli.

Participant	Sticks (long thin)		Pills (small discrete)		Vases (cylindrical)		Rocks (round)		Mushrooms (combination of shapes)	
	Conventional DC	Embellished DC	Conventional DC	Embellished DC	Conventional DC	Embellished DC	Conventional DC	Embellished DC	Conventional DC	Embellished DC
1	1	1	2	2	2	2	2	1	2	2
2		2	6			2		6		1
3	3	0	1			6			10	7
4		5	2	2	4	1		1		4
5		3	5			4		2		6
6		3	2			2				2
7		2	1			2		1		2
8		3	1	3		2		8		6
9		5				6		2		6
10		3	3	1	1	6	1	2		6
11		2	3			2		2		6
12		3	3			4		2	4	3
13		8	1	2		3				10
14		12	11			6			1	6
15		7	5	6	4	15		5		6

For each stimulus, we counted the total number of conventional DCs and embellished DCs the participant produced for each pair of objects (e.g., Stick 1 and Stick 2; Vase 1 and Vase 2).



the stem of a broccoli and a Claw-5 (👉) handshape on the dominant hand on top of the C (👉) handshape to represent the florets (Figure 2). Two-handed DCs do not always contain two different handshapes; e.g., two of the same handshapes (👉) can combine to form a round configuration like a rock (Figure 2).

Adding a tracing movement

Adding movement to a conventional handshape allows signers to capture the object's shape (see Figure 2). This type of sign, which traces the outline of the object in 3D space, has been called a "tracing SASS" (Supalla, 1982), a "contour sign" (Zwitserslood, 2003), "molding," or "sculpting" where the hands "shape a transient sculpture in space" (Müller, 2013; see also Kendon,

2004; Nyst, 2016). For example, a signer moves two Claw-5 (👉) handshapes in and out while going from bottom to top in space in order to sculpt the outline of a vase. At times, signers used an index finger (👉), rather than a classifier handshape, to sketch or draw an object's contour (Mandel, 1977; Müller, 1998; Müller, 2013; Nyst, 2016). Both of these types of tracing DCs can be performed either with one hand or with two identical hands (see Figure 2).

The two embellishing strategies that the signers used in our data to modify their handshapes both mimetically depicted aspects of the stimuli. However, the strategies lent themselves to capturing different features and thus were often used for different stimuli. The signers tended to add movement to depict long, thin objects and cylindrical objects; to add a second handshape to depict small, discrete objects; and to use both strategies (at approximately the same rates) to depict round objects and objects with a combination of shapes. We combined these strategies into a single category, which we called *Embellished DCs*.

In a few cases, there were challenges in distinguishing DCs from lexical signs that may have originated as DCs. For example, the sign for BOTTLE resembles a tracing DC and presumably was derived, at some point, from this spontaneous construction (Cormier et al., 2012). These ambiguous signs were relatively rare ($99/2004 = 0.05$ of all observations) and were excluded from the analyses.

In addition to depicting on the hands, signers also produced movements with their mouths that captured aspects of the objects, often the same aspects captured by the hands (see Sutton-Spence and Boyes Braem, 2001; Sandler, 2009). Using an expanded version of Anderson and Reilly's (1998) coding system for mouth movements (e.g., glosses such as *ps* to indicate cheeks sucking in or *puff* to indicate puffed cheeks), we identified three types of mouth movements: mouthing, lexical mouth components, and iconic mouth movements.

Mouthing

Mouthing lexical words that are borrowed from spoken language (Sutton-Spence and Boyes Braem, 2001; Crasborn et al., 2008); e.g., mouthing the word "bottle" while producing the lexical sign for BOTTLE).

Lexical Mouth Components

Mouth movements that obligatorily co-occur with specific lexical items, but are not derived from speech (Meir and Sandler, 2008; Sandler, 2009); e.g., in Israeli Sign Language, a mouth movement "fa" has to be obligatorily produced with a sign meaning THE-REAL-THING; "fa" has no relation to the words in Hebrew that mean 'the real thing.'

Iconic Mouth Movements

Mouth movements that depict the size and shape of the objects. These movements often capture aspects of the object that are simultaneously captured on the hands; e.g., puffing one's cheeks three times as the hands trace the three bulges of the vase. This category includes mouth movements that Sandler (2009) categorized as adverbial or adjectival modification. However,

in our study, signers rarely used a single adjectival mouth morpheme to modify a nominal sign, as in puffed cheeks used to mean big (Liddell, 1980; Sutton-Spence and Woll, 1999; Crasborn et al., 2008). In our study, signers typically produced sequences of mouth movements (rather than a single movement) to highlight a property of the object (presumably because of the types of objects we presented).

Reliability

A second coder, a hearing signer fluent in ASL, coded 20% of the participants to establish reliability. The coders agreed on 83% of decisions categorizing DCs and 80% of decisions categorizing mouth movements. Coders discussed their disagreements and reached full consensus on the categories.

RESULTS

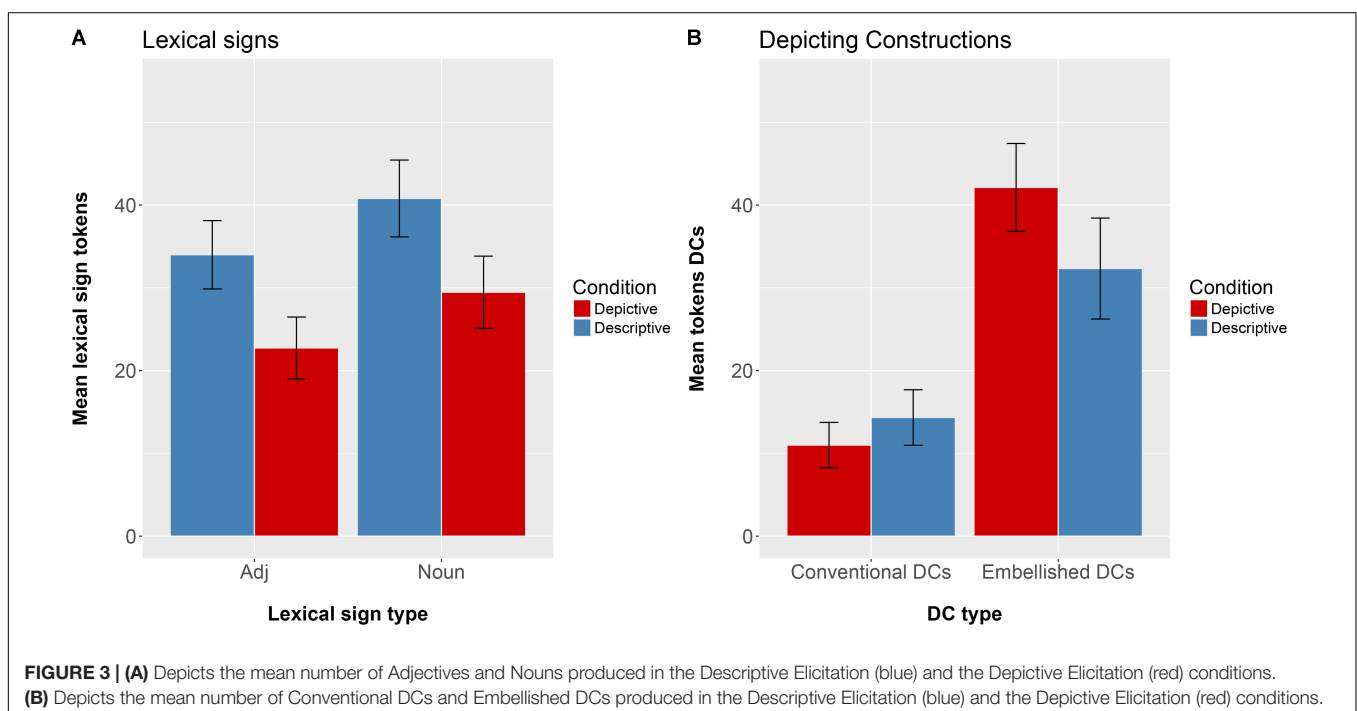
Signs Produced in Depictive vs. Descriptive Elicitation Contexts

Figure 3A presents the mean number of lexical signs (adjectives and nouns) and **Figure 3B** presents the mean number of DCs (conventional and embellished) produced by our participants in the Descriptive Elicitation condition and in the Depictive Elicitation condition.

We first analyzed the patterns of lexical signs produced in Descriptive vs. Depiction conditions. We performed a 2 (Condition: Descriptive vs. Depictive) \times 2 (Word type: Nouns vs. Adjectives) repeated measures ANOVA, using count of lexical signs as the dependent variable. As expected, there was a significant main effect of Condition, where participants produced more lexical items or fingerspelled words, either

nouns (e.g., 'bottle' and 'vase') or adjectives (e.g., 'thin' and 'yellow'), to describe the pairs of objects in the Descriptive Elicitation condition than in the Depictive Elicitation condition [$F(1,14) = 51.13$, $p < 0.0005$]. There was also a significant main effect of Word type, where subjects produced more nouns than adjectives [$F(1,14) = 8.99$, $p < 0.005$]. Finally, there was no significant interaction between Condition and Word type [$F(1,14) = 0.0002$, $p = 0.90$].

Next, we analyzed the patterns of DCs produced in Descriptive vs. Depiction conditions. We performed a 2 (Condition: Descriptive vs. Depiction) \times 2 (Conventional DCs vs. Embellished DCs) repeated measures ANOVA, using count of DCs as the dependent variable. There was a significant main effect of Condition, where subjects produced more DCs in the Depictive Elicitation condition than in the Descriptive Elicitation condition [$F(1,28) = 1.26$, $p < 0.05$]. There was also a significant main effect of DC type, where participants produced more embellished DCs than conventional DCs [$F(1,14) = 22.57$, $p < 0.0005$]. There was also a significant interaction between Condition and DC type [$F(1,28) = 8.54$, $p < 0.005$]. We investigated this interaction further with *post hoc* tests, and found that, at an alpha level of 0.025, signers produced significantly more embellished DCs in the Depictive Elicitation condition than in the Descriptive Elicitation condition ($p < 0.005$). In contrast, the number of conventional DCs that the participants produced did not vary by condition ($p = 0.05$). Embellished DCs have better potential to mimetically capture the size and shape of the objects than conventional DCs. The signers took advantage of this potential and used more embellished DCs in the Depictive condition than in the Descriptive condition. In contrast, they used the same number of conventional DCs in the two



conditions, underscoring the depictive limitations of this type of DC.

Our paradigm was thus successful in eliciting depiction in signers. We focus for the remainder of this paper on DCs produced in the Depictive Elicitation condition.

Conventional vs. Embellished DCs

Table 1 presents the total number of conventional DCs and embellished DCs produced by each participant. Note that all 15 participants produced instances of each type of DC.

We explored whether signers used different depictive strategies depending on the stimulus type. As we saw in **Figure 1**, the objects fell into five different shape categories: (1) long and thin objects (e.g., a stick), (2) small and discrete objects (e.g., pills), (3) cylindrical objects (e.g., a vase), (4) round objects (e.g., a rock), and (5) objects with a combination of shapes (e.g., a mushroom which has both a round part and a long thin part). We selected one representative stimulus from each shape category, and analyzed the total number of conventional vs. embellished DCs that each participant produced (see **Table 2**). Signers used conventional DCs primarily for the small and discrete objects, and embellished DCs for all of the other objects.

Signers also produced a combination of DCs to describe a single object. For example, to describe a rock, a signer first produced a Claw-5 (👉) handshape with her left hand, and then, with her right hand in a B (👉) handshape, traced a curved trajectory on the Claw-5 (👉) and thus produced an embellished DC. She ended by producing a second embellished DC — two Claw-5 (👉) handshapes touching at the fingertips representing the rock as a whole (**Figure 4**, top row). This sequence could be described as a decomposed depiction, where the parts of the object are described sequentially (Sowa and Wachsmuth, 2009). The sequence contains a “frame hold” in which the non-dominant hand remains still while the dominant hand performs the tracing depiction (the embellished DC).

As an example of a combination of embellished DCs used to describe a mushroom, one signer first added a tracing movement to two C (👉) handshapes to depict the stem, and then combined two handshapes—the non-dominant C (👉) handshape depicting the stem and the dominant Claw-5 (👉) hand depicting the mushroom cap (**Figure 4**, bottom row). This is an example of a decomposed depiction where the signer uses two forms of embellished DCs to form an entire image of the object.

Mouth Movements in the Depictive Eliciting Condition

Signers used mouth movements with over half of their lexical signs ($M = 55\%$, $SD = 15\%$). They also used mouth movements with about a third of their Embellished DCs ($M = 36\%$, $SD = 12\%$), but used them with very few of their Conventional DCs ($M = 8\%$, $SD = 5\%$). We found a significant effect of sign type on the number of mouth movements that accompanied the signs [$F(2,28) = 25.01$, $p < 0.0001$]. Lexical signs were significantly more likely to be produced with mouth movements than both types of DCs (embellished DCs, $p < 0.01$; conventional DCs, $p < 0.01$), and embellished DCs were significantly more likely

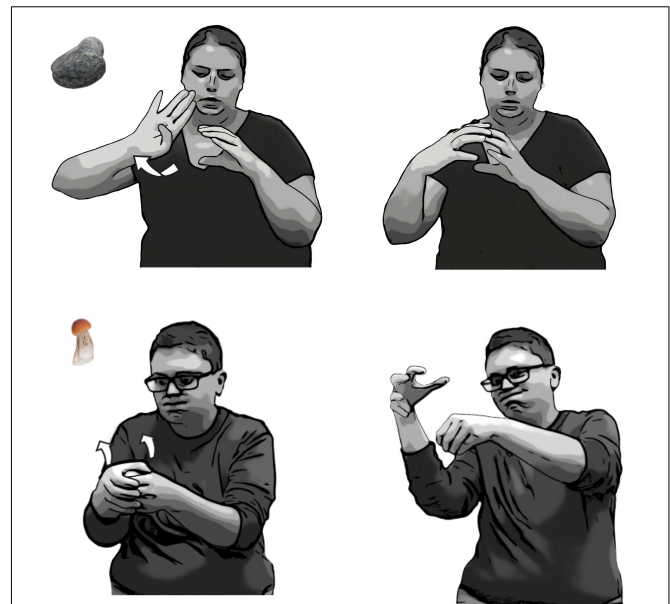


FIGURE 4 | Examples of a combination of embellished DCs. **(Top)** In the image on the left, the non-dominant hand forms a Claw-5 (👉) handshape, while the B (👉) handshape on the dominant hand produces an embellished DC (in which she adds a tracing movement to capture the curvature of the rock). In the right image, the signer ends with another embellished DC (in which she adds a second handshape) to depict the whole shape of the rock. **(Bottom)** To depict a mushroom, the signer first adds a tracing movement to two C (👉) handshapes to depict the stem, and then combines two handshapes, the non-dominant C (👉) handshapes to depict the stem and the dominant Claw-5 (👉) handshape to depict the mushroom cap.

to be produced with mouth movements than conventional DCs ($p < 0.01$).

Moreover, signers used different mouth movements with each sign type. **Figure 5** presents the mean number of tokens of lexical signs, conventional DCs, and embellished DCs that co-occurred with iconic mouth movements (green bars), mouthing (red bars), and lexical mouth movements (blue bars). We ran a 3x3 repeated measures ANOVA (Sign type: Conventional DCs, Embellished DCs, and Lexical Signs) × Mouth Movement type (Iconic mouth movements, Mouthing, Lexical mouth components) using count of signs as the dependent variable. We found a significant main effect of Mouth Movement type [$F(2,28) = 26.45$, $p < 0.0005$], where signers produced significantly more lexical mouthings ($p < 0.005$) and significantly more iconic mouth movements ($p < 0.005$) than lexical mouth components. We also found a significant main effect of Sign type [$F(2,28) = 24.99$, $p < 0.0005$], where signers produced significantly more lexical signs than conventional DCs ($p < 0.005$); and significantly more embellished DCs than conventional DCs ($p < 0.005$).

Importantly, there was a significant interaction between Mouth Movement type and Sign type [$F(4,56) = 61.71$, $p < 0.0005$]. We followed up this interaction with *post hoc* tests at the alpha level of 0.017. Mouthings appeared more often with lexical signs than with either conventional DCs ($p < 0.0005$) or embellished DCs ($p < 0.0005$), and more often

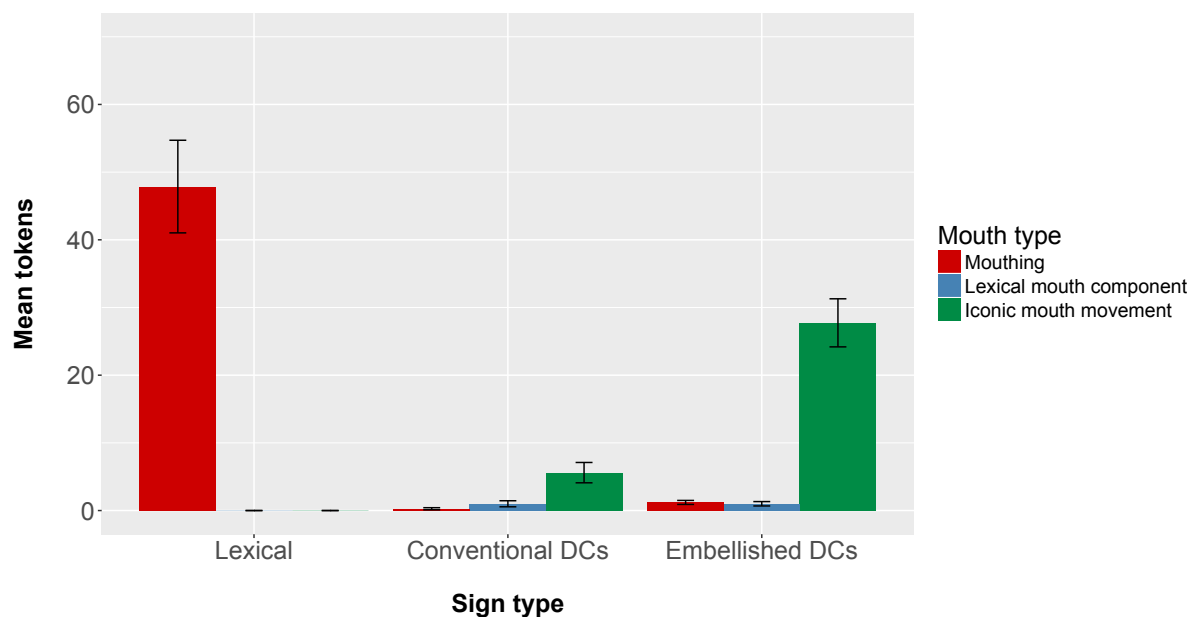


FIGURE 5 | Mean number of Lexical Signs, Conventional DCs, and Embellished DCs produced by participants in the Depictive Elicitation Condition, classified according to whether they co-occurred with Mouthings, Lexical mouth movements, or Iconic mouth movements.

with conventional DCs than embellished DCs ($p < 0.05$). In contrast, iconic mouth movements appeared significantly more often with DCs (both embellished, $p < 0.0005$, and conventional, $p < 0.05$) than with lexical signs, and more often with embellished DCs than with conventional DCs ($p < 0.0005$). Lexical mouth components were rarely produced with any of the three sign types. The fact that iconic mouth movements co-occurred most often with embellished DCs underscores the imagistic aspect of these depictive signs.

Iconic Mouth Movements Produced With Embellished DCs

Signers frequently exploited the same iconic mapping in their iconic mouth movements that they displayed in their embellished DCs. For example, one signer sucked in his cheeks (a *ps* mouth movement), which evokes an imagery of thinness, while at the same time tracing the bottom, thinner part of the vase with two Claw-5 (👉) handshapes (the hands were held close together in space). The mouth then transitioned to *puffed cheeks* while the hands traced the top, wider part of the vase (the distance between the hands increased (Figure 6)². The change from one mouth shape to another is gradual, and is tightly correlated with the changes in the space between the two hands in the embellished DC.

Not only do signers gradiently modify their mouth shapes, transitioning from one mouth shape to another, but they often

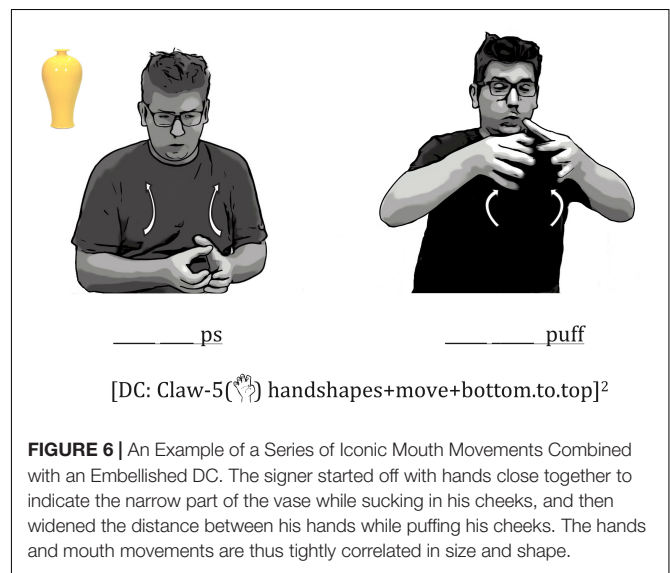
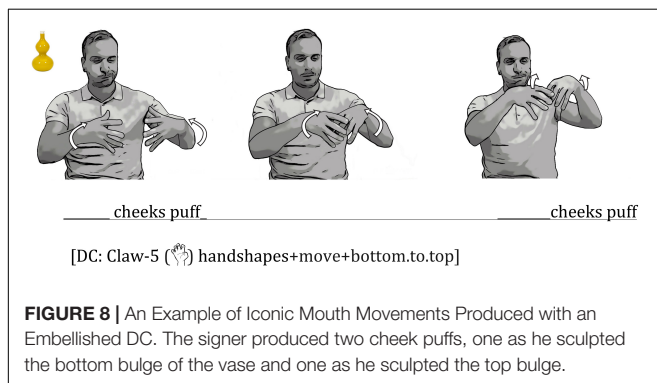
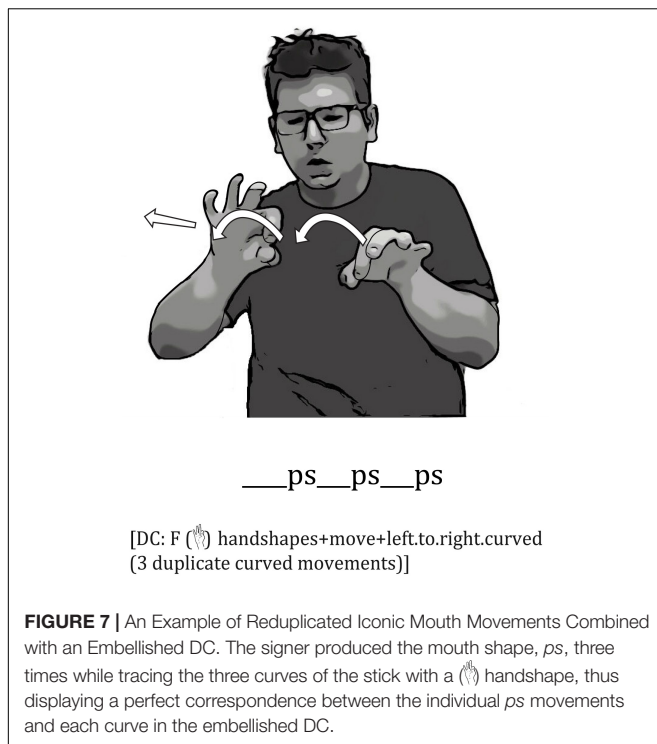


FIGURE 6 | An Example of a Series of Iconic Mouth Movements Combined with an Embellished DC. The signer started off with hands close together to indicate the narrow part of the vase while sucking in his cheeks, and then widened the distance between his hands while puffing his cheeks. The hands and mouth movements are thus tightly correlated in size and shape.

reduplicate the same mouth shape to reflect repeated properties of the object. These mouth reduplications correspond to the spatial reduplications of tracing movements in the embellished DCs. For example, in describing a tree branch with three curves, one signer reduplicated his mouth shape, *ps*, three times as he traced the three curves with his dominant hand in a G (👉) handshape, thus displaying a perfect correspondence between his mouth movements and the hand in his embellished DC (Figure 7).

In another example of reduplicated mouth movements combined with an embellished DC, a signer produced two cheek

²For some figures, we have added information about iconic mouth movements using Anderson and Reilly's (1998) notation system, denoted with underlines. For example, *puff* indicates that the cheeks were puffed out with air, and *ps* indicates that the cheeks were sucked in. There is also information about depicting constructions in brackets, which indicates the handshape being used and the type of movement involved in the production.



puffs while tracing the two round parts of a yellow vase (Figure 8). The signer puffed his cheeks while first tracing the bottom round part of the vase; as he sculpted the middle part of the vase with his hands, he shrunk his cheek puffs; finally, as he traced the top round part of the vase with his hand, he puffed his cheeks again.

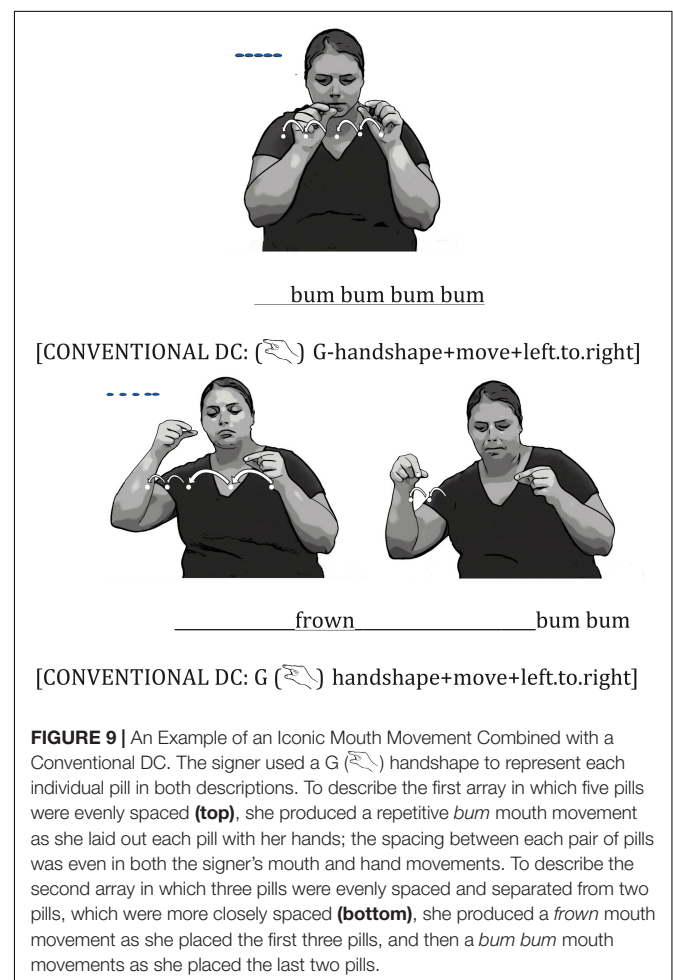
Iconic Mouth Movements Produced With Conventional DCs

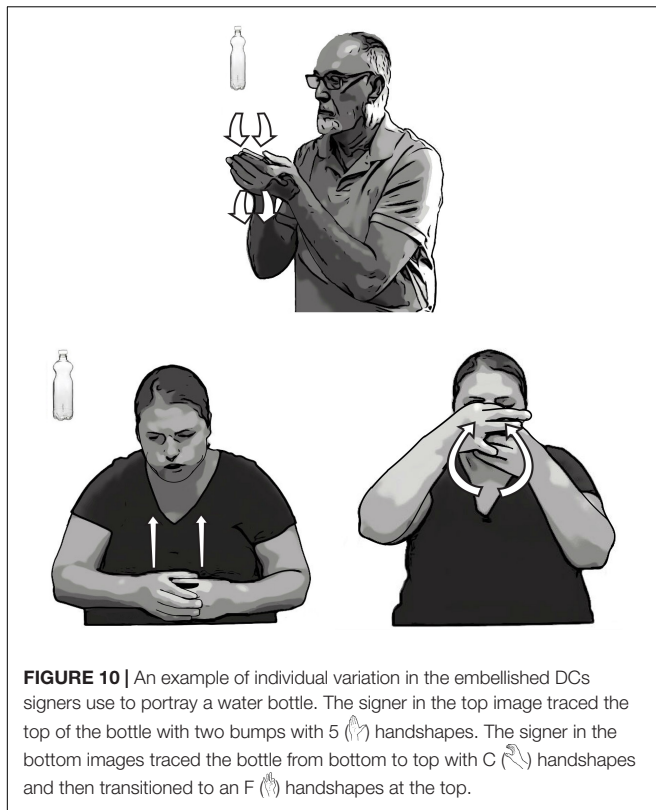
Conventional DCs were often produced to describe the smaller objects, particularly those laid out in distinct arrays. For example, to describe two arrays of pills (see Figure 1), one signer used a conventional DC handshape (*G*) to represent each individual pill, and laid the set of DCs out in space. Note that signing space is used differently in conventional DCs than in embellished DCs. In embellished DCs, space is used to represent the shape of a single object, but in conventional DCs, space is used to represent an arrangement of multiple objects. The signer also

varied her mouth movements to capture the distance between the pills. For the first set of five pills, where the pills were evenly spaced, the signer produced a repetitive *bum* mouth movement as she laid out each pill. For the second set of five pills, where the first three were evenly spaced and separated from the second 2 (which were closely spaced), the signer produced a *frown* mouth movement to capture the relatively wide distance between the pills for the first three pills, and then a repetitive *bum bum* mouth movement to capture the closer distance between the last two pills (see Figure 9). Iconic mouth movements co-vary with hand movements in conventional DCs, as they do in embellished DCs.

Individual Differences in Depicting

Signers varied in the particular movements they produced in their embellished DCs. These variations suggest that embellished DCs were likely to have been created in the moment rather than drawn from a conventional store of movements. Some signers were more veridical to the size and shape of the objects they described, and traced the entire object. Other signers were less specific and captured only the distinguishing features of the objects in their tracings. For example, one signer was attentive to the details of the shape of a water bottle in her description of the bottle and traced its entire contour, depicting the narrower circumference



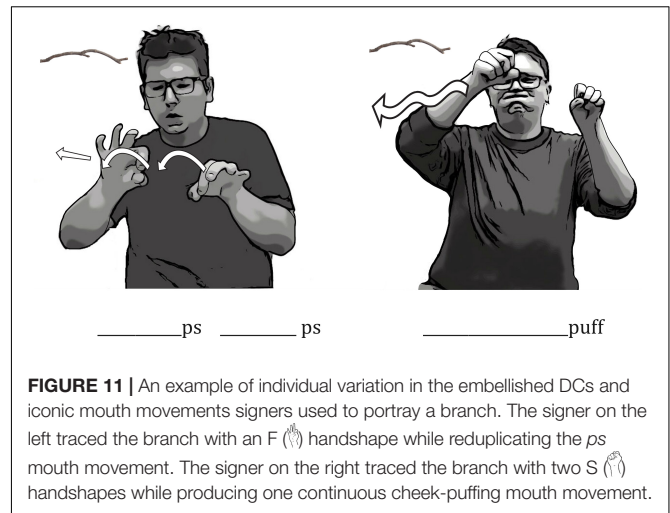


of the bottom half of the bottle and the two bumps on the top (**Figure 10**, bottom row). Another signer did not trace the bottom of the bottle and represented only the two bumps on the top in her tracing DC (**Figure 10**, top). Additionally, these two signers used slightly different handshapes—one signer used 5 (✋) handshapes on both hands (**Figure 10**, top) and the other signer used C (✋) handshapes and then transitioned to F (✋) handshapes as she traced the top of the bottle (**Figure 10**, bottom).

Signers also varied in how they combined the two handshapes in their embellished DCs. For example, in depicting a rock, some signers produced a combination of DCs by holding a Claw-5 (✋) handshape on the left, and then using a B (✋) handshape on the right hand to trace the curvature (**Figure 4**). Other signers would instead hold a Claw-5 (✋) handshape on the left, and then place another Claw-5 (✋) handshape on the top and not use any tracing movement.

The variation across individuals in hand tracings and hand combinations was paralleled by variation in iconic mouth movements. To describe a stick, one signer traced the branch with an F (✋) handshape and reduplicated the *ps* mouth shape (**Figure 11**). To describe the same stick, another signer used two S (✋) handshapes that were slightly open to trace the curve of the stick, and produced one continuous mouth movement by puffing her cheeks and bottom lip.

In contrast to the variability found in embellished DCs and iconic mouth movements, there was (not surprisingly) less variability across signers in the conventional DCs. Signers tended to use the same handshape to represent a particular shape, for



example, a G (✋) handshape for small and discrete items, and a C (✋) handshape to represent cylindrical objects, each with a hold movement. Moreover, they used relatively few mouth movements with conventional DCs and, when they did use a mouth movement with this type of sign, they tended to draw from a relatively small set of mouth movements (e.g., usually single cheek puffs).

DISCUSSION

Signers use multiple strategies to depict the size and shape of objects. When it is difficult to succinctly distinguish between two objects using lexical labels, signers resort to depiction. They recruit conventional and embellished DCs, and combine each type with iconic mouth movements to imagistically capture the sizes and shapes of the objects they are describing. However, embellished DCs occurred more often in Depictive contexts than in Descriptive contexts, whereas conventional DCs occurred equally often in the two contexts. Moreover, embellished DCs co-occurred more frequently with iconic mouth movements than conventional DCs, and may be more tightly integrated, temporally and spatially, with these mouth movements than conventional DCs. Taken together, these findings suggest that embellished DCs share properties with co-speech gestures in that both are imagistic and spontaneously created.

Depicting Constructions vs. Lexical Signs

Depicting constructions are a heavily used resource for signers to describe the size and shape of objects. Signers often choose to depict rather than use available lexical signs, such as LONG, TALL, or MIDDLE. Interestingly, when asked to distinguish between the same pairs of objects, speakers often use a litany of adjectives and invoke specific scenarios (e.g., “a vase that you can put a flower in”; Lu and Goldin-Meadow, 2017). In one example (6), an English speaker said the following to distinguish between the two yellow vases in **Figure 1**:

- (6) “One of the objects in this one is a yellow... Looks like a vase that you can put a flower in. Um it’s like it gets *slimmer* as you go toward the *bottom* whereas the other object could also be a flower vase, there are like two different bumps in the *middle* and at the *bottom*.”

The speaker also produced many iconic co-speech gestures along with his many adjectives, *slim*, *middle*, and *bottom*, to describe different aspects of the vase. Signers rarely used the sign equivalents of these adjectives, even though they are available in ASL. They relied entirely on depictive devices to convey size and shape.

Why might it be easier to depict in sign rather than use descriptive lexical items, and what prompts signers to rely on these depictive devices? We can imagine several factors that could lead to heavy use of depiction in sign: (1) It may be particularly efficient to use DCs, which can encode two characteristics of an object (e.g., width and height) simultaneously within a single construction; encoding the same information lexically would require several signs (Quinto-Pozos, 2007; Sevcikova Sehyr et al., 2018). (2) Many lexical signs in ASL have an iconic base, and these iconic properties might clash with the meaning of the objects being described (cf. Meir, 2010); using a depictive device would circumvent this potential difficulty. For example, the sign THIN involves using two pinky fingers that first contact each other and then move in opposite directions vertically in space. The vertical movement of this lexical form nicely captures objects that are thin and upright, such as a candle. Signers are likely to use this lexical sign in this case. However, this form is a less good rendition of objects that are thin and *horizontal*, such as the sticks in our study. Signers may therefore be less inclined to use the lexical sign THIN, which is produced vertically in space, to describe a horizontal stick. Instead, they turn to an embellished DC, tracing a G (👉) handshape horizontally in space. When lexical signs do not map neatly onto their referents, signers may choose to depict using embellished DCs so that they can be faithful to the iconic mapping. Future work is needed test this hypothesis by exploring whether lexical signs are more likely to be used than depictions when they can be fully mapped onto the form of their referents. (3) ASL signers and English speakers have different lexicons (Quinto-Pozos, 2007; Sevcikova Sehyr et al., 2018), and some of the stimuli in our study were difficult to label with a single lexical sign in ASL (e.g., *knob* of the French press; *florets* of the broccoli; *tip* of the screwdriver). Signers may need to resort to depiction, using embellished DCs in particular, to describe items that English speakers can reference with lexical labels (there are, of course, items that do not have ready labels in English, and we predict that English-speakers will rely on depiction for these items).

Embedded Depiction in Signed and Spoken Languages

Depicting constructions play an important role within a signed utterances in that, if they are removed, significant information is lost. DCs can be analyzed as a constituent within a clause-like unit (Ferrara and Johnston, 2014; Hodge and Johnston, 2014), just as embedded depictions can be analyzed as linguistic units in spoken

languages (Clark, 2016). In the *Bartok* example presented earlier (2), the pianist starts his sentence with, “he does not play,” and then depicts what the pianist doesn’t do by playing a short Mozart passage, a depiction that functions as a noun phrase.

In our data, signers often began by first naming the object (e.g., using the lexical sign for vase) and then following that description with a spontaneously created depiction (e.g., tracing the contour of the vase; using two hands to indicate the configuration of the vase). These depictions thus function like adjectival predicates. Similar structures can be found in speech. For example, a speaker begins by describing the vase (“there is a vase”), and then switches into depiction by tracing the two bumps of the vase in a co-speech gesture; gestures of this sort are often accompanied by sound effects (in this case, *bum bum*), which seem to function like iconic mouth movements. As in sign, this depiction serves as an adjectival predicate, and the shape information is not found anywhere else within the clause.

These depictions in speech and sign contain gestural materials in the sense that the forms are not conventional, and are unlikely to have been drawn from a lexicon of words or signs. Nevertheless, the forms often take on the full weight of expressing information about the size and shape of an object (which is not conveyed anywhere else in the utterance). The gestural materials work together with the linguistic materials to form a composite utterance (Enfield, 2009; Ferrara and Johnston, 2014).

Relevant phenomena, such as constructed action in sign and demonstrations in speech, have both been analyzed as predicates, each providing action information that is not found anywhere else in the sentence. In the following examples, the first in sign (7) and the second in speech (8), the constructed action functions as a verb phrase:

- (7) BOY (CA: THE BOY LOOKS INTO THE HOLE) Hodge and Johnston (2014)
 (8) I got out of the car, and I just (*demonstration of turning around and bumping his head on an invisible telephone pole*) Clark and Gerrig (1990)

In the clause-like unit in (7), the actor is the boy, who is identified with the lexical sign BOY. However, there is no lexical verb describing the boy’s action of looking through the hole, other than the enactment in the constructed action. The signer uses his torso, head, and face to represent the boy, and enacts the looking process. In the clause-like unit in (8), the gestural demonstration also functions as a verb-like predicate, conveying action information that is not found in the speech. In our data, signers often use depiction as the only source of information about the size and shape of objects. These depictions thus appear to function as an adjectival predicate within the clause.

Linguistic Constraints on Depictions

Handshapes in DCs have been analyzed as morphemes (Supalla, 1982, 1986) and we agree with this analysis. Indeed, both conventional and embellished DCs appear to be categorical (and conventional) in that the signers in our study always used the same handshape to refer to a particular type of object, for example, a G (👉) handshape to refer to small and discrete

items (such as pills) or the Claw-5 (🐾) handshape to refer to round objects (such as rocks). Overall, there was very little variability in the handshape signers used to depict different shapes of objects, which suggests that handshape is conventional in these constructions (Frishberg, 1975; Kegl and Wilbur, 1976; Zwitserlood, 2003).

But handshape does not provide the full meaning of the object in embellished DCs (Zwitserlood, 2003). Signers added movement to the handshape to capture the referent's shape. The movements they added were gradient (rather than categorical) and signers varied in the shapes they sculpted with their moving hands (Emmorey and Herzig, 2003; Liddell, 2003; Schembri et al., 2005). Nevertheless, there were still linguistic constraints on these embellished DCs. Signers could trace the contours of an object using either an index finger (which carves out a 2D representation) or another conventional handshape (which carves out a 3D representation). For the most part, signers chose the conventional handshape that captured a feature of the object. For example, signers used either a B (👉) or Claw-5 (🐾) handshape to trace the contour of a vase, but did not use a less appropriate handshape such as a G (👉) handshape. As another example, signers often used a C (👉) handshape to represent the stem of a mushroom and combined it with a Claw-5 (🐾) handshape to represent the cap; they rarely used other handshapes to represent this object. Handshape in an embellished DC is conventionally determined by the to-be-described object, following the same constraints that arise in conventional DCs. The key difference is that, in embellished (but not conventional) DCs, the handshape is modified with gradient movement or with another handshape to further specify the object's shape.

Iconic mouth movements also provide shape information about objects, and may reflect an interaction between motoric and linguistic constraints. The mouth is not as free as the hands to convey shape through three-dimensional space. As a result, there are a limited number of mouth shapes available to signers that can function as conventional, adjectival mouth morphemes (Liddell, 1980; Sutton-Spence and Woll, 1999; Crasborn et al., 2008). For example, most signers puffed their cheeks when describing a vase. However, signers displayed variation in how they modified the *cheek puff* to capture the contour of the vase: one signer transitioned from a cheek puff to *ps*; another puffed multiple times to describe the same vase. The modifications overlaid on top of the mouth movements thus appear to be idiosyncratic and, in this sense, gestural.

Embellished DCs and modified iconic mouth movements can thus both be analyzed as categorical forms that are gradiently modified. These productions are comparable to analog speech in spoken language, where categorical forms are modified in a meaningful and iconic way (Shintel et al., 2006). For example, speech can be modified analogically by elongating the vowel "o" in the word *long*: "It was a *loooooong* time." The conventionalized, categorical form "long" is modified gradiently to add emphasis to the meaning; it was not just a long time, but a *really* long time. However, there are constraints on how words can be modified analogically. For example, one cannot elongate other parts of the word and thus cannot say *lllllong* or *longngngng*; the vowel is a more likely candidate for modification

than the consonant (Okrent, 2002). This constraint parallels the constraints that signers face when they use a particular handshape or mouth movement to construct a depiction.

Depicting Constructions Share Properties With Spoken Ideophones

A special class of sensory words in spoken language—mimetics or ideophones found in African languages like Siwu or Japanese (Dingemanse and Akita, 2016)—may be a good analog to DCs in sign language. Like DCs, these spoken devices are iconic words that are borderline linguistic. They are flexible and amenable to gradient modification via reduplication or vowel lengthening. For example, the ideophone, *gat(-to)*, meaning a 'rattling sound,' can be reduplicated to *gagagagagagat-to* to depict the reverberating sounds of debris falling (note that the morpheme within the ideophone, *gat*, can itself be iconic). Interestingly, ideophones frequently co-occur and are tightly integrated with iconic co-speech gestures, and often depict the same meaning as those gestures—they "perform the same role of depicting sensory imagery, albeit in different modalities and therefore also with different affordances for iconicity" (Dingemanse and Akita, 2016; see also Kita, 1997). For example, a speaker talks about an incoming wave using the ideophone, *zorot(-te)*, meaning 'one after another in line,' and reduplicates it, *zorozorot-te* while producing a time-aligned, reduplicated iconic gesture—he swings his arms from right to left twice.

Depicting constructions are similar to ideophones in two ways. First, categorical handshapes in signs are combined with non-discrete movement, just as conventional morphemes in ideophones are imagistically and analogically reduplicated, resulting in high expressivity. Second, DCs are tightly coupled with iconic mouth movements, just as ideophones are tightly coupled with co-speech gestures. If there is a spoken language that uses ideophones to describe the shape of objects, we predict that speakers of this language would use ideophones cohesively with iconic co-speech gesture.

Conventions in the Practice of Depicting

We raise one last point with respect to depiction—although aspects of the form of depiction may not be entirely conventional, the practice of depicting may be conventional. The degree of conventionalization of a construction can be analyzed on two levels: conventionalization of the form itself (which we have discussed), and conventionalization of how the form is used (Okrent, 2002). Evidence from an emerging sign language shows that some linguistic devices can be conventionalized even before phonology emerges (Meir et al., 2010). In Al-Sayyid Bedouin Sign Language (ABSL), signers often use two signs to label a single object. For example, to identify a pen, signers often sign WRITE, followed by a conventional DC ('SASS classifier') referring to a thin object. Even though ABSL signers often choose different aspects of the pen to highlight in their conventional DCs, they are remarkably similar in how they order their two signs (DCs occupy the second position in 90% of instances).

In our data, depictions arise frequently, and we speculate, in a predictable way. For example, Enfield (2009) notes that there

may be conventions with respect to when and how co-speech gestures are used (see also Lu and Goldin-Meadow, 2017). In fact, listeners can use the speaker's eye gaze toward gesture space to predict when the speaker is going to produce a tracing gesture. Future work is needed to determine whether there is systematicity in when and how signers use embellished depictions, how embellished depictions relate to other constituents within the clause, and whether these depictions are foreshadowed by other cues.

How Necessary Are DCs to Convey the Full Communicative Message?

Depicting constructions contribute significant meaning to linguistic utterance. Indeed, as in other depictive phenomena in spoken and signed languages (Clark and Gerrig, 1990; Liddell, 2003; Quinto-Pozos, 2007; Hodge and Johnston, 2014; Clark, 2016), the message would be incomplete without DCs. Signers may show a strong preference for depictive devices over lexical items in some communicative contexts simply because depictions can often provide more depth and accuracy in portraying a referent than lexical signs. In future work, our goal is to elicit judgments of signed utterances that contain either depiction or lexical items, and to assess how much information can be gleaned from these two types of utterances, how obligatory depictive devices are, and whether depictive devices provide richer meanings than descriptive devices.

We have demonstrated several ways in which depictive devices play an important role in conveying meaning in sign language. Importantly, this phenomenon is not unique to sign languages. Yet depiction tends to be relegated to the margins of language sciences and ignored in standard models of language (Liddell, 2003; Clark, 2016; Dingemanse and Akita, 2016). We have shown that signers, at times, will choose depiction over description as their primary communicative strategy, thus signaling the importance of depiction in discourse. The interesting question is whether depiction is just as important in spoken languages as it is in signed languages, a question that can only be answered by

exploring depiction under comparable circumstances in speech and sign.

ETHICS STATEMENT

All procedures for all studies reported were approved under University of Chicago IRB #15-1678. All participants provided written informed consent to participate in this study, and those identified in the images of the manuscript provided written informed consent for their publication.

AUTHOR CONTRIBUTIONS

JL and SG-M designed the research. JL performed the research and analyzed the data. JL and SG-M wrote the paper.

FUNDING

This work was supported by the National Science Foundation Graduate Research Fellowship Program Grant No. DGE-1144082 (to JL), the Institute of Education Sciences and United States Department of Education Grant No. R305B140048 (to S. Raudenbush in support of JL), and the Center for Gesture, Sign, and Language Grant (to JL and SG-M).

ACKNOWLEDGMENTS

The authors thank Zena Levan for assistance with coding and illustrations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01276/full#supplementary-material>

REFERENCES

- Anderson, D. E., and Reilly, J. (1998). PAH! the acquisition of adverbials in ASL. *Sign Lang. Linguist.* 1, 117–142.
- Brentari, D., and Padden, C. A. (2001). "Native and foreign vocabulary in American Sign Language: a lexicon with multiple origins," in *Foreign Vocabulary: A Cross-linguistic Investigation of Word Formation*, ed. D. Brentari (Mahwah, NJ: Lawrence Erlbaum Associates), 87–119.
- Chawla, P., and Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *J. Exp. Soc. Psychol.* 30, 580–601. doi: 10.1006/jesp.1994.1027
- Clark, H. H. (2016). Depicting as a method of communication. *Psychol. Rev.* 123, 324–347. doi: 10.1037/rev0000026
- Clark, H. H., and Gerrig, R. J. (1990). Quotations as demonstrations. *Language* 66, 764–805. doi: 10.1037/rev0000026
- Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behav. Res. Methods Instrum. Comput.* 25, 257–271. doi: 10.3758/BF03204507
- Cormier, K., Quinto-Pozos, D., Sevcikova, Z., and Schembri, A. (2012). Lexicalisation and de-lexicalisation processes in sign languages: comparing depicting constructions and viewpoint gestures. *Lang. Commun.* 32, 329–348. doi: 10.1016/j.langcom.2012.09.004
- Cormier, K., Smith, S., and Sevcikova, Z. (2015). Rethinking constructed action. *Sign Lang. Linguist.* 18, 167–204. doi: 10.1075/sll.18.2.01cor
- Crasborn, O., Van Der Kooij, E., and Waters, D. (2008). Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Lang. Linguist.* 11, 45–67. doi: 10.1075/sll.11.1.04cra
- Dingemanse, M., and Akita, K. (2016). An inverse relation between expressiveness and grammatical integration?: on the morphosyntactic typology of ideophones, with special reference to. *J. Linguist.* 53, 501–532. doi: 10.1017/S002222671600030X
- Duncan, S. (2005). Gesture in signing: a case study from Taiwan sign language. *Lang. Linguist.* 6, 279–318.
- Emmorey, K., and Herzig, M. (2003). "Categorical versus gradient properties of classifier constructions in ASL," in *Perspectives on Classifier Constructions in Sign Languages*, ed. K. Emmorey (Mahwah, NJ: Lawrence Erlbaum Associates), 221–246.

- Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511576737
- Engberg-Pederson, E. (1993). *Space in Danish Sign Language: The Semantics and Morphosyntax of the Use of Space in a Visual Language*. Hamburg: Signum Press.
- Ferrara, L., and Johnston, T. (2014). Elaborating who's what: a study of constructed action and clause structure in Auslan (Australian sign language). *Aust. J. Linguist.* 34, 193–215. doi: 10.1080/07268602.2014.887405
- Frishberg, N. (1975). Arbitrariness and iconicity: historical change in American sign language. *Language* 51, 696–719. doi: 10.2307/412894
- Hodge, G., and Johnston, T. (2014). Points, depictions, gestures and enactment: partly lexical and non-lexical sign as core elements of single clause-like units in Auslan (Australian Sign Language). *Aust. J. Linguist.* 34, 262–291. doi: 10.1080/07268602.2014.887408
- Hostetter, A. B., Alibali, M. W., and Kita, S. (2007). I see it in my hands' eye: representational gestures reflect conceptual demands. *Lang. Cogn. Process.* 22, 313–336. doi: 10.1080/01690960600632812
- Johnston, T., and Schembri, A. (2010). "Variation, lexicalization and grammaticalization in signed languages," in *Sourds et Langue des Signes: Normes et Variation. Langage et Société*, Vol. 131, eds B. Garcia and M. Derycke (Paris: Editions de la Maison des sciences de l'homme), 5–17.
- Kästner, L., and Newen, A. (2017). Good things come in threes: communicative acts comprise linguistic, imagistic, and modifying components. commentary on: S. Goldin-Meadow & D. Brentari, gesture, sign and language: the coming of age of sign language and gesture studies. *Behav. Brain Sci.* 40, 1–82. doi: 10.1017/S0140525X15001247
- Kegl, J., and Wilbur, R. B. (1976). Where does structure stop and style begin? Syntax, morphology, and phonology vs. stylistic variation in American sign language. *Chic. Linguist. Soc.* 12, 376–397.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572
- Kita, S. (1997). Two-dimensional semantic analysis of Japanese mimetics. *J. Linguist.* 35, 379–415. doi: 10.1515/ling.1997.35.2.379
- Liddell, S. K. (1980). *American Sign Language Syntax*. Hague: Mouton.
- Liddell, S. K. (2003). *Grammar, Gesture and Meaning in American Sign Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511615054
- Lu, J., and Goldin-Meadow, S. (2017). "Signers and gesturers use different depictive strategies along with language in communication," in *Proceedings of the 11th International Symposium on Iconicity in Language and Literature*, Brighton.
- Mandel, M. (1977). "Iconic devices in American Sign Language," in *On the Other Hand: New Perspectives on American Sign Language*, ed. L. Friedman (New York, NY: Academic Press), 57–107.
- Meir, I. (2010). Iconicity and metaphor: constraints on metaphorical extension of iconic forms. *Language* 86, 865–896. doi: 10.1353/lan.2010.0044
- Meir, I., Aronoff, M., Sandler, W., and Padden, C. (2010). "Sign languages and compounding," in *Cross-Disciplinary Issues in Compounding*, eds S. Scalise and I. Vogel (Amsterdam: John Benjamins Publishing), 301–322. doi: 10.1075/cilt.311.23mei
- Meir, I., and Sandler, W. (2008). *A Language in Space: The story of Israeli Sign Language*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Metzger, M. (1995). "Constructed dialogue and constructed action in American Sign Language," in *Sociolinguistics in Deaf Communities*, ed. C. Lucas (Washington, DC: Gallaudet University Press), 255–271.
- Müller, C. (1998). *Rede begleitende Gesten. Kulturgeschichte — Theorie — Sprachvergleich*. Berlin: Berlin Verlag Arno Spitz.
- Müller, C. (2013). "Gestural modes of representation as techniques of depiction," in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*, eds D. McNeill and J. Bressemer (Berlin: De Gruyter Mouton), 1687–1701.
- Nyst, V. (2016). The depiction of size and shape in gestures accompanying object descriptions in Anyi (Côte d'Ivoire) and in Dutch (The Netherlands). *Gesture* 15, 156–191. doi: 10.1075/gest.15.2.02nys
- Okrent, A. (2002). "A modality-free notion of gesture and how it can help us with the morpheme vs. gesture question in sign language linguistics (Or at least give us some criteria to work with)," in *Modality and Structure in Signed and Spoken Languages*, eds R. P. Meier, K. Cormier, and D. Quinto-Pozos (Cambridge: Cambridge University Press), 175–198.
- Quinto-Pozos, D. (2007). Can constructed action be considered obligatory? *Lingua* 117, 1285–1314. doi: 10.1016/j.lingua.2005.12.003
- Sandler, W. (2009). Symbiotic symbolization by hand and mouth in sign language. *Semiotica* 2009, 241–275. doi: 10.1515/semi.2009.035
- Schembri, A., Jones, C., and Burnham, D. (2005). Comparing action gestures and classifier verbs of motion: evidence from Australian sign language, Taiwan sign language, and nonsigners' gestures without speech. *J. Deaf Stud. Deaf Educ.* 10, 272–290. doi: 10.1093/deafed/eni029
- Sevcikova Sehyr, Z., Nicodemus, B., Petrich, J., and Emmorey, K. (2018). Referring strategies in American Sign Language and English (with co-speech gesture): The role of modality in referring to non-nameable objects. *Appl. Psycholinguist* 1–27. doi: 10.1017/S0142716418000061
- Shintel, H., Nusbaum, H. C., and Okrent, A. (2006). Analog acoustic expression in speech communication. *J. Mem. Lang.* 55, 167–177. doi: 10.1016/j.jml.2006.03.002
- Slobin, D. I., Hoiting, N., Kuntze, M., Lindert, R., Weinberg, A., Pyers, J., et al. (2003). "A cognitive/functional perspective on the acquisition of 'classifiers,'" in *Perspectives on Classifier Constructions in Sign Languages*, ed. K. Emmorey (Mahwah, NJ: Laurence Erlbaum)
- Sowa, T., and Wachsmuth, I. (2009). "A computational model for the representation and processing of shape in coverbal iconic gestures," in *Spatial Language and Dialogue*, eds K. R. Coventry, T. Tenbrink, and J. A. Bateman (Oxford: Oxford University Press), 132–146.
- Streeck, J. (2008). Depicting by gesture. *Gesture* 8, 285–301. doi: 10.1075/gest.8.3.02str
- Supalla, T. (1982). *Structure and Acquisition of Verbs of Motion and Location in American Sign Language*. San Diego, CA: University of California.
- Supalla, T. (1986). "The Classifier System in American Sign Language," in *Proceedings of the Symposium on Categorisation and Noun Classification*, Amsterdam, 181–214. doi: 10.1075/tsl.7.13sup
- Sutton-Spence, R., and Boyes Braem, P. (2001). *The Hands are the Head of the Mouth: The Mouth as Articulator in Sign Languages*. Hamburg: Signum Press.
- Sutton-Spence, R., and Woll, B. (1999). *The Linguistics of British Sign Language: An Introduction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139167048
- Zwitzerlood, I. (2003). *Classifying Hand Configurations in Nederlandse Gebarentaal (Sign Language of the Netherlands)*. Doctoral dissertation, Utrecht University Repository, Utrecht.
- Zwitzerlood, I. (2012). "Classifiers," in *Sign Language: An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 158–186.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer RS and handling Editor declared their shared affiliation.

Copyright © 2018 Lu and Goldin-Meadow. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Metaphor in Sign Languages

Irit Meir^{1,2†} and Ariel Cohen^{3*}

¹ Department of Hebrew Language, University of Haifa, Haifa, Israel, ² Department of Communication Sciences and Disorders, University of Haifa, Haifa, Israel, ³ Department of Foreign Literatures and Linguistics, Ben-Gurion University of the Negev, Beersheba, Israel

Metaphor abounds in both sign and spoken languages. However, in sign languages, languages in the visual-manual modality, metaphors work a bit differently than they do in spoken languages. In this paper we explore some of the ways in which metaphors in sign languages differ from metaphors in spoken languages. We address three differences: (a) Some metaphors are very common in spoken languages yet are infelicitous in sign languages; (b) Body-part terms are possible in very specific types of metaphors in sign languages, but are not so restricted in spoken languages; (c) Similes in some sign languages are dispreferred in predicative positions in which metaphors are fine, in contrast to spoken languages where both can appear in these environments. We argue that these differences can be explained by two seemingly unrelated principles: the Double Mapping Constraint (Meir, 2010), which accounts for the interaction between metaphor and iconicity in languages, and Croft's (2003) constraint regarding the autonomy and dependency of elements in metaphorical constructions. We further argue that the study of metaphor in the signed modality offers novel insights concerning the nature of metaphor in general, and the role of figurative speech in language.

Keywords: metaphor, simile, iconicity, inhibition, Double Mapping Constraint, autonomous and dependent elements

INTRODUCTION

Metaphor, the use of an item from one semantic domain in a different semantic domain in order to characterize the latter in terms of the former, is pervasive in human language and thought. Though it is often regarded as a poetic device used in figurative language to create special poetic effects, works on metaphor in the past several decades have demonstrated that metaphors are used in everyday use of language, and not only in language but in thought and action as well (Lakoff and Johnson, 1980). In fact, we cannot avoid using metaphors; all we need is to look and we will catch metaphors in many everyday utterances (note that *look* and *catch* are used metaphorically here). Since our potential experiences are infinite, yet the lexicon of any language is finite, the use of metaphor is a powerful way to refer to new situations by using the existing linguistic means that we have (e.g., *surfing* the internet, a computer *mouse*, a *spaceship*).

Furthermore, metaphor is not restricted to language; it is used in other domains of human cognition as well, such as mathematics (Nunez, 2008), visual art (Kennedy, 1982, 2008; Forceville, 2008), graphics, and music (Zbikowski, 2008).

Natural languages come in two modalities—spoken and signed. Both types of languages develop naturally in human communities, shaped by the special characteristics of the human brain and human capacity for language, by human cognition and by the communicative needs and constraints of human communities. The languages produced in the two modalities have

OPEN ACCESS

Edited by:

Marianne Gullberg,
Lund University, Sweden

Reviewed by:

Alan Cienki,
VU University Amsterdam,
Netherlands
Gary Lupyan,
University of Wisconsin-Madison,
United States

*Correspondence:

Ariel Cohen
aric@bgu.ac.il

[†] Deceased

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 14 January 2018

Accepted: 31 May 2018

Published: 26 June 2018

Citation:

Meir I and Cohen A (2018) Metaphor
in Sign Languages.
Front. Psychol. 9:1025.
doi: 10.3389/fpsyg.2018.01025

many important properties in common, in their linguistic structures, processes, constraints and communicative functions (Sandler and Lillo-Martin, 2006). Since metaphor seems to be such a basic and pervasive cognitive process, we would expect to find it in both types of languages.

Yet the two types of languages differ markedly in their physical characteristics, and these physical characteristics entail some important linguistic differences between the two modalities. For example, in sign languages, both the articulations of the hands and their relation to space are directly perceivable, unlike those of the vocal tract, whose articulations are perceivable only indirectly, via the acoustic patterns created by air passing through different vocal tract configurations. Sign languages also fully exploit the existence of the two hands – phonologically, lexically, and at higher levels of structure. These two identical articulators can behave independently, and have no parallel in speech (see e.g., Sandler, 1993, 2006, 2013; Liddell, 2003; Crasborn, 2012).

These modality differences result in structural differences as well (see e.g., Meier et al., 2002). For example, sign languages exhibit more simultaneous structure on all linguistic levels (Sandler and Lillo-Martin, 2006; Vermeerbergen et al., 2007), while spoken languages show a tendency toward sequential structures. In addition, iconicity is more pervasive on all linguistic levels in sign languages than in spoken languages (Johnston and Schembri, 1999; Aronoff et al., 2005; Meir, 2010; Lepic et al., 2016).



What about metaphor? Would we expect languages in the two modalities to behave differently with respect to metaphorical expressions? Why should we, or why shouldn't we, expect such differences? As we pointed out above, metaphors can be found in visual forms of communication and art. Therefore, we would expect to find them in visual languages too. However, metaphors in visual systems may work differently than metaphors in spoken languages. Kennedy (2008, p. 455) points out that a metaphor such as *a burning passion* works well in spoken language, but a picture of a burning person misses the point entirely. He points out that additional physical details, that cannot be avoided in a picture, are often extraneous and distracting. Kennedy further notices (ibid., p. 458) that in pictures one cannot distinguish between a metaphor (*my daughter is an angel*) and a simile (*my daughter is like an angel*)¹.

Sign languages are both visual systems and linguistic systems. We might expect metaphor to work in a similar way in languages in general, building on the properties shared by all human languages. Yet if modality does play a role in shaping metaphors, as suggested above, then metaphors may work differently in the two types of languages.

Research on metaphors in sign languages reveals that metaphor is abundant in these languages. The seminal work of Wilcox (2000) and Taub (2001) on metaphor in sign languages showed that metaphorical mapping plays a central role in creating signs, especially signs for abstract concepts. Moreover, they show that the types of mappings found in ASL are those

mentioned by Lakoff and Johnson (1980) as forming the basis for conceptual metaphors in spoken languages, such as: GOOD IS UP, THE FUTURE IS AHEAD, INTIMACY IS PROXIMITY, COMMUNICATION IS SENDING, UNDERSTANDING IS GRASPING, and many more. More recently, Roush (2016) examined eleven sub-mappings of location event-structure metaphors, which are claimed to be universal in spoken languages, and found that all of them are exhibited in signs from the ASL lexicon. These studies provide strong support for prevalence of conceptual metaphor in human language, regardless of modality.

Important insights can be obtained from the study of metaphorical gestures (e.g., Cienki and Müller, 2008), although gestures, unlike signs, usually co-occur with speech. While McNeill (1992) distinguishes between iconic and metaphorical gestures, Cienki and Müller (2008) argue that metaphorical gestures are in fact iconic. The iconic nature of gestures is particularly important, as it “affords different potentials than aural/oral expression does” (Müller and Cienki, 2009, p. 322).

Wilcox (2000) and Taub (2001) focus on the interaction of metaphor and iconicity in the structure of signs. They show how the different phonological components of a sign – its hand configuration, location and movement – can represent iconically some of the meaning components of that sign, and then can be metaphorically mapped to an abstract concept in a different semantic domain. For example, the sign EAT in Israeli Sign Language (ISL) has the form of a  handshape, moving in a repeated movement toward the signer's mouth. The form of the sign iconically represents holding a small object (by the  handshape), and putting it into the agent's mouth (Figure 1). When the hand performs the same movement toward the temple, the sign means LEARN, represented iconically as the action of putting something inside one's head (Figure 2). In this sign, the iconic representation is mapped onto the abstract domain of mental activities, which is characterized by putting objects (ideas, information, pieces of knowledge) into a container (the head).

Taub (2001) elaborated on the relationship between iconicity and metaphor, and suggested an explicit model capturing the relationship between the two. Specifically, she suggests that the creation of an iconic sign is a process of mapping elements of form to elements of meaning. And the creation of an iconic-metaphorical sign is shaped by double mapping: an iconic mapping from form to meaning components, and a metaphorical mapping from the meaning components (the source domain of the metaphor) to the target domain of the metaphor.

As we demonstrate in this paper, our own work on metaphors in sign languages, based on Taub's model, shows that indeed metaphors in sign languages work a bit differently from spoken languages. First, some metaphors that are very common in spoken languages cannot receive a metaphorical interpretation in the signed modality (Meir, 2010). Second, while in spoken languages a word or an expression that are interpreted metaphorically have the same form as their non-metaphorical counterpart, in sign languages often metaphorical use of sign also involves slight changes in the form of the sign (Cohen and Meir, 2014). Furthermore, in sign languages similes are often less favored than their metaphorical counterparts, in linguistic environments in which both are acceptable in spoken languages.

¹Forceville (1996, 2005) draws a distinction between pictorial metaphors and what he calls 'similes,' but by this he just means cases where the source and target domains are visually presented separately.

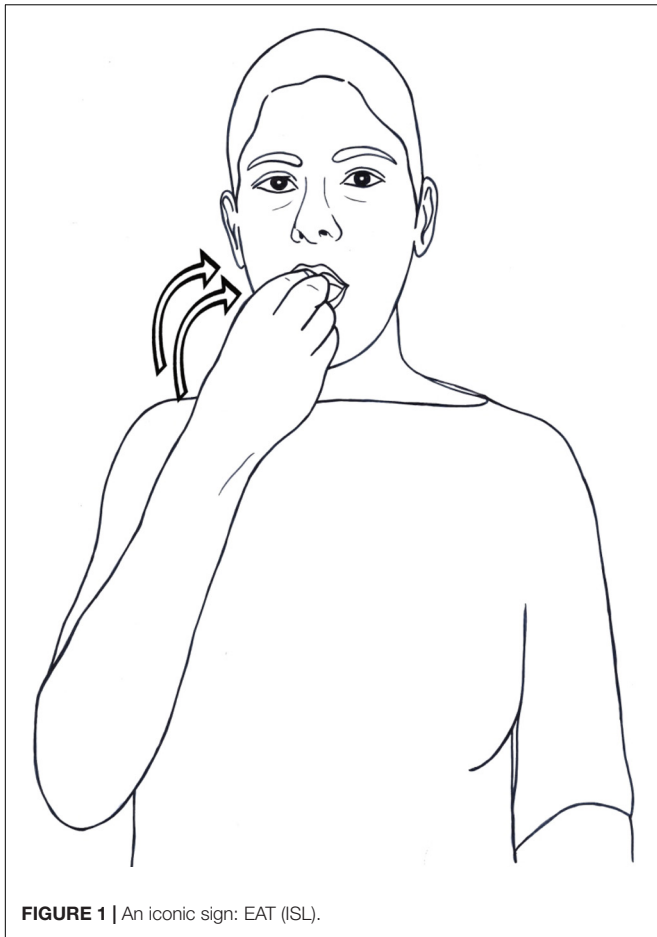


FIGURE 1 | An iconic sign: EAT (ISL).

In the current paper we explore some of the ways in which metaphors in sign languages differ from metaphors in spoken languages, and suggest explanations to these differences. We further argue that the study of metaphor in the signed modality offers novel insights concerning the nature of metaphor in general, and the role of figurative expressions in language.

In what follows, we describe and account for three types of differences between metaphors in sign languages and in spoken languages. In Section “The Interaction Between Iconicity and Metaphor,” we focus on the interaction between iconicity and metaphor, showing that the iconicity of signs constrains the metaphorical interpretations they can get. We introduce the Double Mapping Constraint (Meir, 2010), and suggest that it can explain the differences between languages in the two modalities, as well as shed light on the predication nature of metaphor. Body-part terms, a common source for metaphors in spoken languages, show variable behavior concerning participation in metaphorical expressions in sign languages. We argue that this variable behavior can be explained by the interaction between the DMC and Croft’s (2003) constraint (see section “The Body in Metaphors”). Croft’s constraint is also used to explain two additional, seemingly unrelated, differences between sign and spoken languages, namely that in sign languages similes are often

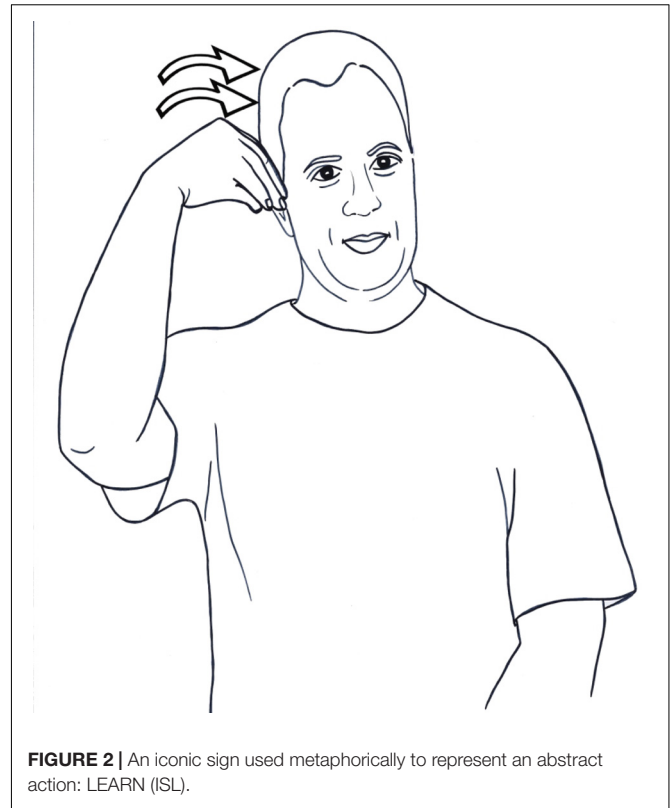


FIGURE 2 | An iconic sign used metaphorically to represent an abstract action: LEARN (ISL).

dispreferred while their metaphor counterparts are acceptable, and the fact that often metaphorical signs have a slightly different form than their non-metaphorical counterparts (see section “Similes and Metaphors in ISL”). We conclude by describing two intriguing difference between the use of metaphors in signed vs. spoken languages, to which we do not yet have an explanation, and which we leave for future research (see section “Conclusions and Future Work”).

A word on methodology is in order here. The data presented in this paper are based on consultation with three ISL native signer, and an ASL native signer, as well as some informal discussions with a few more fluent ISL signers. Though there are differences and variation among signers regarding specific possible and impossible metaphors and figurative expressions in ISL, there was general agreement regarding the data presented here.

THE INTERACTION BETWEEN ICONICITY AND METAPHOR

The Double-Mapping Constraint


Metaphor involves mapping between source and target domains. However, not any such mapping is acceptable. For example, Lakoff (1990) formulates what he calls the Invariance Hypothesis, according to which metaphorical mappings between source and target domain are partial, and the portion of the source domain which is mapped preserves the image schematic structure of the source domain that is topologically consistent with the structure

of the target domain. Thus, metaphors only map structure from the source domain that is compatible with the target domain.

Meir (2010) notes that in sign languages metaphorical mapping is further constrained, as some expressions that receive a metaphorical interpretation in spoken languages cannot be so interpreted in sign languages. For example, (1–3) normally do not mean that the house/acid/car literally ate all my savings/the metal/gas, but rather that these substances were consumed by the event that took place.

- (1) The house ate up all my savings.
- (2) The acid ate through the metal.
- (3) My car eats gas.

However, this metaphorical interpretation of the verb *to eat* is unavailable when these sentences are translated to sign languages, such as American or Israeli Sign Languages. Meir attributes the unavailability of metaphorical interpretation to the iconicity of the sign EAT in these languages, whose form represents putting something into the agent’s mouth (Figure 1 above). She suggests that the iconicity of this sign clashes with the shifts in meaning that take place in these metaphorical extensions. This explanation is based on Taub’s (2001) model that both iconicity and metaphors are built on mappings of two domains: form and meaning in iconicity, source domain and target domain in metaphors. Iconic signs that undergo metaphoric extension are therefore subject to both mappings.

Yet, this double mapping is not always available. When the two mappings do not preserve the same structural correspondence, Meir (2010) argues that the metaphorical extension is blocked. This line of explanation accounts for the impossibility of using the ISL sign EAT in the above expressions. The meaning of ‘eat’ is ‘to put (food) in the mouth, chew if necessary, and swallow.’ That is, the food is consumed as a result of the eating event. But the consumption of the food is not represented iconically in the form of the sign. The form of the sign iconically represents holding a small object (by the  handshape), and putting it into the agent’s mouth (represented by the movement of the hand toward the signer’s mouth). Each of the formational components of the sign (its handshape, location and movement) corresponds to a specific meaning component of the event of eating, as is shown in the left and middle columns of Table 1.

But the metaphorical use of *eat* in the above sentence profiles the consumption: *The house ate up my savings* means that the house consumed my savings as the agent consumes the food in an eating event. The metaphorical mapping between the two

domains is presented by the middle and right columns of Table 1. The two mappings, the iconic mapping and the metaphoric mapping, do not match, as can be seen from Table 1 (Meir, 2010, p. 879).

The meaning component that is active in the metaphorical mapping, the consumption, is not encoded by the iconic form of the sign. And the meaning components of the iconic mapping – the mouth, manipulating an object, putting into mouth – are bleached in the metaphor. The mismatch in the double mappings of the verb EAT and its intended metaphorical interpretation suggests that there is some kind of interaction between the iconic form of a sign and the kinds of metaphorical extensions it can undergo. Specifically, the iconic form of a concept and its metaphorical extension cannot profile different aspects of that concept. This is captured in the following constraint (Meir, 2010, p. 879):

The Double-Mapping Constraint (DMC): A metaphorical mapping of an iconic form should preserve the structural correspondences of the iconic mapping. Double-mapping should be structure-preserving.



The DMC can account for other metaphors that are possible in many spoken languages but not in sign languages, such as *Time flies*, *He climbed the ladder of success*, *the project took off*. In each of these expressions, the concept undergoing metaphorical extension is represented in ISL and ASL by an iconic sign, whose form highlights aspects of the meaning that should be bleached in the metaphor. In FLY (Figure 3), the hands represent the flapping of the wings, a meaning component irrelevant for the metaphor. The metaphor profiles the speed of motion, which is not represented by the form of the sign. Similarly, the form of the ISL sign CLIMB highlights the manner of motion (moving by grasping the wrings of the ladder in an alternating fashion) rather than the upward movement intended as the basis for the metaphoric interpretation; and the form of the ISL sign TAKE-OFF highlights (by its  handshape) the instrument performing the action (an airplane), which is irrelevant for the metaphor.

TABLE 1 | Double mapping for EAT and ‘consuming is eating.’

Iconic mapping		Metaphorical mapping
Articulators	Source	Target
 handshape	Holding an object (food)	X
Mouth	Mouth of eater	X
Inward movement	Putting food into mouth	X
X	Consumption of food	Consumption of object

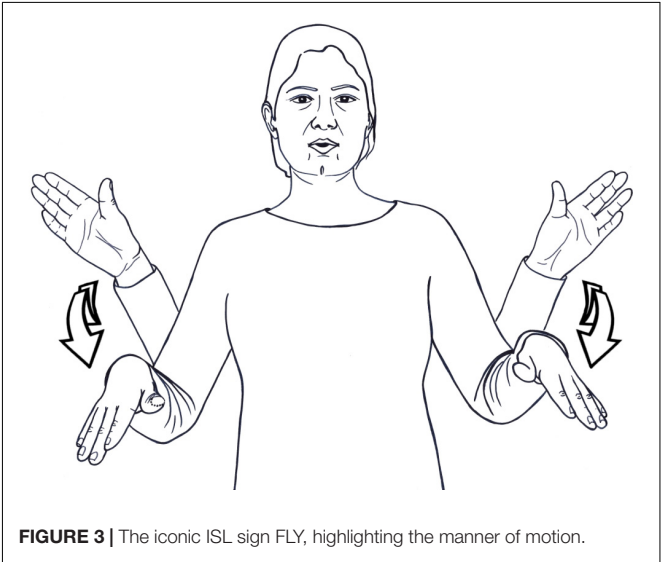


FIGURE 3 | The iconic ISL sign FLY, highlighting the manner of motion.

The Source for the DMC: Inhibition in Metaphor and Iconicity

The DMC suggests that iconicity interacts with metaphor in an interesting way: it restricts the possibility of an iconic sign to be used or interpreted metaphorically, if the property which is iconically represented in the form of the sign is not the property which the metaphor is based on. But what is the source for this restriction? Why does iconicity interfere with metaphorical extension of a sign? We attribute this interference to another property of iconic expressions, the fact that iconicity cannot be inhibited. Yet metaphorical interpretation requires the inhibition of certain properties of the word. It is the tension between these two factors that will feature prominently in our explanation of the DMC. Inhibition, then, is crucial to our suggestion, to which we turn in this section. We first look at the role of inhibition in metaphoric interpretation, and then at its interaction with iconicity.

Metaphor and Inhibition

Intuitively, in order to interpret a metaphoric statement such as (4), we need to inhibit the literal properties of the sun, such as being very massive or very hot or 150 million kilometers from Earth; we keep only the properties that are relevant to the interpretation of the metaphor. The notion that metaphor interpretation requires inhibition has received some experimental confirmation.

Glucksberg et al. (2001) show that properties that are not relevant to the metaphorical interpretation are negatively primed, i.e., inhibited. Subjects were asked to judge the acceptability of target sentences following either metaphors or literal statements. For example, a sentence like (6) was judged following either the literal (5a) or the metaphorical (5b).

- (4) Juliette is the sun.
- (5) (a) The hammerhead is a shark.
- (b) My lawyer is a shark.
- (6) Geese can swim.

Glucksberg et al. (2001) found that (6) took longer to judge when preceded by (5b) than when preceded by (5a). Their explanation is that the word *shark* normally primes the property *swim*, facilitating the interpretation of a sentence containing it. However, if *shark* is interpreted metaphorically, this property is not primed. This is why (6), which contains the word *swim*, takes longer to judge when following the metaphorical (5b) than when following the literal (5a).

Fernandez (2007) extended these findings, by showing that the irrelevant property is not simply not primed, but actually inhibited. Using a lexical decision task, Fernandez showed that words that are related to the literal meaning of the metaphor actually took longer to judge than words that were not related at all.

For example, consider (7) and (8): in both of them, a target word follows a sentence. The target word in (7b), *skin*, is not related to any of the words of (7), hence it is not primed. In contrast, the target word in (8b), *animal*, is related to the word *zoo* appearing in (8a). But note that *zoo* is used metaphorically in (8a),

and the interesting result is that the judgment of (8b) is actually *slower* than the judgment of (7b)! This result demonstrates that the literal meaning of *zoo* is actually actively inhibited, not merely not primed. Interestingly, the effect occurs only after 1500 ms, which is consistent with the fact that inhibition takes time.

- (7) (a) Wisdom teeth are troublemakers.
- (b) Skin
- (8) (a) State schools are zoos.
- (b) Animal

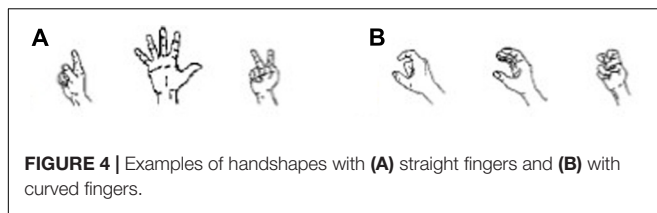
Langdon et al. (2002) provide evidence for the inhibition hypothesis from a different direction: the behavior of schizophrenic patients. In particular, they studied both the ability of these patients to inhibit irrelevant information and their interpretation of metaphors, and found the following correlation: “the better the patients were at suppressing prepotent inappropriate information... the more likely they were to recognize appropriate uses of metaphorical speech.”

Iconicity and Inhibition

Thompson et al. (2010) found that iconic signs are much harder to inhibit than non-iconic ones, even in tasks that require no access to meaning. They asked deaf signers of British Sign Language (BSL) to make a phonological decision: to decide whether BSL signs, presented in video clips, were produced with a handshape with straight or curved fingers (see **Figure 4**). The signs were both iconic and non-iconic, but importantly, the iconicity of the signs was irrelevant for the task, as the task did not involve access to the meaning or meaning components of the signs. Thompson et al. (2010) found that iconic signs led to slower reaction times and more errors in the participants' responses. They suggest that meaning is activated automatically for highly iconic signs, because of the closer form-meaning mapping in these signs². This automatic activation of meaning interfered with the task because it provided information that could not be inhibited yet was irrelevant to the task at hand. It seems, then, that iconicity cannot be ignored, even when it is irrelevant.

Another possible inhibitory effect of iconicity was found by Baus et al. (2013). The tasks in this study did involve meaning, as bilingual (ASL-English) signers were asked to translate signs (iconic and non-iconic) from ASL to English and from English to ASL, or to determine whether a given ASL sign and a given English word match in meaning. The findings show that iconicity interfered with the performance of fluent ASL-English bilinguals: their responses to the ASL-into-English translation task and the matching task were significantly slower for iconic signs than for non-iconic ones. These results are surprising. In the Thompson et al.'s (2010) study described above, iconicity seemed to interfere with the task because it caused automatic access to meaning, which was irrelevant to the phonological task in that study. Yet in the translation task, faster access to meaning is expected to speed translation for iconic signs. The authors suggest that

²A similar explanation in a different theoretical framework is suggested by Emmorey (2014), who regards iconic representations as structured mapping between two mental representations. She suggests that structure-mapping sometimes cannot be avoided, and that iconic mappings are automatically available to signers.



maybe the iconicity of the signs “forced” the participants to use a specific translation strategy that slowed down performance. In order to translate a word, an association must be formed between the lexical systems of the source and target languages (*word-word association*), or the associations can be formed through the conceptual systems (*conceptually mediated translation*). The authors tentatively suggest that “the imagistic or sensory-motor properties of the iconic signs induced these signs to be translated via conceptual mediation, which slowed translation times” (Baus et al., 2013, p. 269). An explanation along these lines supports the hypothesis that iconic properties of signs cannot be inhibited.

Putting it Together: The Source for the DMC

It seems, then, that iconicity and metaphorical interpretation play a constant tug-of-war game. Metaphorical interpretation requires the inhibition of some aspects of the literal meaning of the word, in particular, those aspects that are irrelevant for the metaphorical reading. Iconic aspects of signs, together with the meaning components they are associated with, on the other hand, cannot be inhibited. They are too salient in the form of the sign to ignore. If the metaphorical reading requires the inhibition of those meaning components that are iconically present in the form of the sign, the metaphoric interpretation is not available. Hence the source for the DMC is the competing and opposing forces that iconicity and metaphor require: inhibition of meanings vs. the impossibility of inhibition of these meanings.³ We now turn to a specific type of source domain for metaphor that is affected by the DMC in an interesting way, namely body-part terms.

THE BODY IN METAPHORS

The Problem

Words denoting body-parts are a rich source for metaphorical use in spoken languages, especially for expressing spatial and containment relations (‘the foot of the hill’), part-whole relations (‘the mouth of the river’), and more abstract relations (‘the heart of the problem’). In ISL and other sign languages, such metaphors are completely absent. They are also impossible; signers we’ve consulted with affirm that that they would never use body-part signs in such contexts. Yet sign languages use body-part signs productively in compound-like constructions, such as the following ISL examples: HEAD+STOP ‘to have a blackout’ (Figure 5),

HEAD+FALL ‘to faint,’ HEAD+COGWHEELS ‘to think deeply,’ ‘EYE+SHARP’ ‘to discern visually,’ MOUTH+SMEAR ‘to mislead (by talking),’ HEAD+EMPTY ‘doesn’t understand anything.’ These constructions are common in various sign languages, such as ISL, ASL, British Sign Language (BSL) and others⁴. In ISL we found about 70 constructions of this type (termed *sense compounds* in Aronoff et al., 2005). They are also productive; signers use body-part terms with other words to create novel expressions. The equivalent English constructions are predicate-argument constructions (e.g., *My head is empty*, *His eyes are sharp*) or predicating-modifier constructions (an *empty-headed person*, *sharp-eyed*).

Why do sign languages allow metaphors involving body-part signs in these constructions but not in relational constructions, while spoken languages allow both? What is special about body-part terms in sign languages that makes them more constrained in terms of the metaphorical extensions they can undergo? The answer to this puzzle involves both the DMC, and a constraint suggested by Croft (2003) regarding the autonomy and dependency of elements in a metaphorical construction. We turn now to introduce Croft’s constraint, then return to offer an explanation of the behavior of body-part signs in metaphors in sign languages.

Croft’s (2003) Constraint

Croft (2003) addresses the issue of what drives listeners to interpret a construction metaphorically rather than literally. For example, in a sentence such as *Denmark shot down the Maastricht treaty* (ibid., 162), how does the listener know that the sentence is about politics rather than about war? *Denmark* and *the Maastricht treaty* are entities in the domain of politics,⁵ while *shoot down* is an action that belongs to the domain of war. Why is *shoot down* interpreted as a political action rather than interpreting *Denmark* and *the Maastricht treaty* as belonging to war? And why not interpret the sentence literally? Croft argues that what drives the metaphorical interpretation is “the conceptual unity of domain: all of the elements of a syntactic unit must be interpreted in a single domain.” (ibid., 162). If the literal interpretation provides different semantic domains, the sentence is not rejected as semantically incoherent. Rather, the listener attempts to interpret some of the elements figuratively, as belonging to the same semantic domain as the other elements in that sentence (ibid., 195).

Yet which element of the unit will be interpreted metaphorically? Here Croft draws on Langacker (1987, 1989, 1991, 2002) distinction between autonomous and dependent elements. Langacker notices that in most grammatical combinations, one notion is relatively autonomous, while the other is relatively dependent in the sense that it presupposes the autonomous element as part of its internal structure or interpretation (Langacker, 2002, p. 122). In the phrase *a tall*

³This conclusion has implications, beyond sign languages, on the general meaning of metaphor. Specifically, it favors the view that metaphor expresses predication rather than categorization (Cohen and Meir, 2015).

⁴See Aronoff et al. (2005); Meir and Sandler (2008) for ISL; Brennan (1990); Sutton-Spence and Woll (1999) for BSL; Carol Padden, p.c. for ASL; Meir et al. (2012) for ABSL.

⁵In fact, *Denmark* is in the domain of geography, and belongs to politics only after it is interpreted metonymically. Although Croft does discuss metonymy, we will not deal with it here.



FIGURE 5 | Compound-like constructions that include body part in ISL: HEAD+STOP 'to have a blackout.'

man, *man* is autonomous, since one can conceive of a man without considering his height; while *tall* is dependent, since its meaning is dependent on the conceptualization of an entity to which a quality of tallness can be attributed (Sullivan, 2009, p. 3). When considering predicative elements such as verbs, adjectives, or adverbs vs. nominal arguments, it is usually the case that the latter are autonomous while the former are dependent.

Croft suggests that this distinction is relevant for figurative interpretation of language. In particular, in metaphor, Croft observes that the dependent element is interpreted metaphorically, while the autonomous elements are interpreted non-metaphorically, and signal the target domain. In the sentence above, *Denmark* and *the Maastricht treaty* are autonomous, while *shoot down* is dependent, as its meaning is elaborated by the two nominal phrases. Therefore, the two nominal phrases are interpreted non-metaphorically, and they indicate the target domain (politics) onto which the dependent element should be mapped. The verb, the dependent element, receives a metaphorical interpretation: its meaning is mapped from the domain of war (its source, or literal domain) to the domain of politics (the target domain).

To take another example, in the sentence *My heart broke*, the noun *heart* and the verb *broke* belong to two semantic domains: *heart* belongs to the domain of emotions (by metonymy, the heart is the location of emotions) while *break* belongs to the domain of solid objects. Domain unity requires both elements to be interpreted as belonging to the same domain. Since *break* is dependent while *heart* is autonomous, it is *break* that is interpreted metaphorically. *Heart* signals the target domain of emotion, while *break*, whose source domain is that of solid objects, is mapped to the domain of emotions, receiving a metaphorical interpretation.

Evidence for this constraint comes from psychological experiments. Gentner and France (1988) found that subjects prefer to generate metaphorical interpretations for verbs rather than nouns. For example, they prefer an interpretation of (9)

where the lizard basked in the sun, rather than an interpretation where the person who looks like a lizard worshipped.

(9) The lizard worshipped.

There is also evidence from corpus studies. Huang (1994) found that in both English and Chinese, metaphorically interpreted elements are predominantly adjectives and verbs, not nouns. Sullivan (2007, 2009), in a corpus study of metaphorical expressions, tested Croft's constraint. She analyzed 2415 metaphorical constructions of six different types. Her findings indicate that Croft's predictions are borne out in each of these six constructions. Two of these constructions belong to the constructions relevant for the sign language data, presented above, to which we turn in the next section.

What is the explanation for Croft's constraint? Croft does not propose one, but we believe it follows from the nature of metaphor. There is a debate concerning what the interpretation of metaphor involves⁶. The prevailing views can be roughly classified into two camps: Class inclusion and Predication.

Consider the following example:

(10) Businesses are dictatorships.

According to class inclusion theories of metaphor, the interpretation of (10) involves the construction of an *ad hoc* superordinate class—*dictatorship*^{*}. This class plausibly contains organizations and communities that are managed non-consensually and punitively by one person. Sentence (10) is then taken to mean that the class *business* is a member of this *ad-hoc* superordinate class.

According to predication theories of metaphor, the interpretation of (10) is different. It assumes that there is a set of relevant properties associated with dictatorships: be

⁶See Chiappe and Kennedy (2001), Glucksberg (2001, 2008, 2011), Bowdle and Gentner (2005), Glucksberg and Haught (2006a,b), Jones and Estes (2006), and Utsumi (2007, 2011), *inter alia*.

a form of government, be ruled by one person, be ruled non-consensually, regulate many aspects of the lives of their members, employ political propaganda, use terror and violence, etc. One such property, *P*, is selected. Sentence (10) then means that business have property *P*. For example, in a given context the selected property may be *be ruled by one person*. Then (10) means that businesses are run by a single person. In this paper we assume the predication theory of metaphor (cf. Cohen and Meir, 2015). One argument for the predication view is that it makes possible a natural explanation of Croft's constraint, which would otherwise be an unmotivated stipulation. The explanation is as follows. An intuition that goes as far back as Plato and Aristotle is that a sentence is divided into subject and predicate, where the subject is typically nominal and the predicate is typically verbal or adjectival⁷. It follows that the prototypical predicative categories are verbs and adjectives, rather than nouns (although all three have the same logical type: properties of individuals). We therefore expect verbs and adjectives to be preferred in metaphorical interpretation, which is precisely what is described by Croft's constraint.

Possible and Impossible Body-Part Metaphorical Extensions in Sign Languages

We turn now to the participation of body-part terms in metaphors in sign languages. Let us address first the question of why body-part terms in sign languages are more constrained than their spoken language equivalents. Again, the key to that question is their form. Body-part terms in sign languages usually take the form of pointing to the relevant body part. The signs for EYE, NOSE, EAR in ISL involve a pointing handshape (☞) to the relevant organs. The sign for HEAD is a ☞ handshape that touches the temple; FACE involves a circle movement of the ☞ handshape around the face; HEART is a ☞ hand that touches the location of the heart, and so on. In all these signs, the actual body part serves as the place of articulation of the sign, and is highlighted by the movement of the hand toward it.

The salience of the actual body part in the sign is also what constrains its use in metaphors. The foot of the hill is not really a foot; it is the lowest part of the body, the one that makes contact with the ground, which is what it has in common with the feet of a human body. But it doesn't have toes, it is not connected to a leg, and it doesn't come in pairs. In spoken languages, the metaphorical use is built on the resemblance of the spatial relations between the foot and the body it is part of, abstracting away from the actual form of the human vs. geographical foot. In sign languages, the actual form of the organ is there as part of the form of the sign, and is highlighted in the sign. It cannot be inhibited, and its actual form cannot be ignored. The metaphorical mapping is therefore incongruent with the iconicity of the sign, violating the DMC, and is consequently blocked.

Yet body-part terms in sign languages seem to be absent from one type of construction, and possible in another

type. This differential behavior can be explained by looking at the relationship between the different components of each construction in terms of their relative dependency. In constructions such as *the mouth of the river*, *the foot of the hill* (Sullivan's 'prepositional phrase constructions'), there is a part-whole relationship between the body-part and the noun in the PP, designating a geographical area. The element denoting the part is the dependent element, since its conception is dependent on the conceptualization of the whole that it is part of. It is impossible to conceive of a mouth without referring to body that it is part of (in our case, river). The entity is relatively autonomous, as it is possible to conceive of a river (or of any body) without referring to specific sub-parts of it (Croft, 2003). According to Croft's constraint, the NP denoting the entity, as the autonomous element, is interpreted literally and signals the target domain (geographical areas), while the body-part, as the dependent element, should receive a metaphorical interpretation. However, in sign languages this is not possible, as pointed out above: the form of signs denoting body-parts highlights the actual body-part. The metaphorical mapping is therefore incongruent with the iconicity of the sign, violating the DMC, and is consequently blocked.

The metaphorical expressions HEAD+EMPTY, HEART+BLACK, exhibit a different pattern of autonomous-dependent relationship between its components. These belong to what Sullivan (2007, 2009) calls 'predicating modifier constructions' (*an empty head*), or to 'predicate-argument constructions' (*your head is empty*). In both cases, the predicating element is conceptually dependent, since its interpretation needs to make reference to an entity to which the relevant property can be attributed. The body-part is autonomous, signaling, by metonymy, the target domain of the metaphor: mental activities or emotions (the head is the site for mental activities, the heart the site of emotions). Since the body-part does not receive metaphorical interpretation, it is not subject to the DMC, and is not blocked by it. Therefore, such constructions are possible in sign languages.

We conclude that body-part signs are indeed excluded from being used metaphorically in sign languages because of their form. But they can be part of a metaphorical construction where they function as the autonomous element, denoting the target domain. The interaction of the DMC with Croft's constraint explains how body-part signs, and iconic signs in general, can participate in metaphors.

SIMILES AND METAPHORS IN ISL

The Distribution of Similes vs. Metaphors in ISL

The phenomena described in the previous sections indicate that metaphorical use is more constrained in sign languages than in spoken languages, which we attributed to the constraining effect of iconicity on metaphor. Since iconicity is much more prevalent in sign languages, this restricting effect is more noticeable in these languages than in spoken languages.

⁷Of course, the subject needs not be a separate phrase or even word, and may be incorporated into the predicate, as it is in many polysynthetic languages.

We now turn to another phenomenon where ISL exhibits a more restricted use of figurative language compared to English and Hebrew: the use of similes. Similes are figures of speech that involve comparison between two things of different kinds, in order to characterize one term by the other. In that, they resemble metaphors. However, in similes, the comparison is made explicit, by using words such as *like*, *as*: *My lawyer is like a shark, He works like a mule*. Importantly, in many linguistic structures in spoken languages, similes and metaphors can be both used, as in (11):

(11) John is (like) a snake.

The relationship between similes and metaphors has been studied extensively for millennia. Starting with Aristotle, many scholars (e.g., Bergmann, 1979; Miller, 1993; van Genabith, 2001) argue that a metaphor is an (elliptical) simile. Yet, others argue that metaphors and similes differ in kind. How do we decide this question? We may try to find languages where similes are allowed but not metaphors, or vice versa. It would seem that if such a language is attested, this would indicate that metaphors cannot be reduced to similes. It turns out that the study of sign languages provides us with such a language, but not with the expected outcome.

To the best of our knowledge, the use of similes vs. metaphors has not been studied in sign languages. It might be expected that since in similes the comparison is explicit, iconicity will not play such a restrictive role regarding their use. On the other hand, since in both metaphors and similes, the characteristics that are profiled by the comparison do not necessarily coincide with those profiled by the iconic form of a sign, similes may show very similar behavior and distribution to that of metaphors. To our surprise, when we started looking at the distribution of metaphors and similes in ISL, we found out that in some environments, such as predicative or adverbial positions, similes are often not possible or dispreferred, where metaphors are possible or preferred, as exemplified in (12–14):

- (12) JOHN (*LIKE) SNAKE 'John is (*like) a snake.'
 (13) MARY WORK (?LIKE) MULE 'Mary works (?like) a mule.'
 (14) KIM STRONG (?LIKE) OX 'Kim is strong (?as) an ox.'

Note that in the English sentences that correspond to (13) and (14), not only is the simile form possible, but it is, in fact, mandatory: the metaphor form, that is, the form without *like/as*, is unacceptable.

What is the source of these differences between the two types of languages? To answer this question, let us again consider the English sentence (11). In this sentence, under either its metaphor or simile form, the noun *snake* receives a figurative interpretation. But doesn't this fact violate Croft's constraint? Nouns are regarded as relatively autonomous, and should not receive figurative interpretation according to Croft. However, Croft (2003) speculates that the noun can be construed as dependent after all: "While there appears to be no general principle by means of which we can say that the metaphorically interpreted noun is... dependent... it seems to be a not unreasonable hypothesis... and should be investigated further"

(p. 194). But sign languages allow another option, namely a shift in the lexical category of the noun.

Categorical Reinterpretation of Figurative Signs

Sign languages in general show more flexibility regarding lexical categorical distinctions, in that words in many sign languages are often multicategorical and can be interpreted as nouns, verbs or adjectives (Meir, 2012 and references therein).

"We also find a substantial amount of systematic ambiguity or vagueness in many sign languages. For instance, in Indo-Pakistani Sign Language (IPSL) many signs tend to have rather general meanings that are narrowed down by the context of the utterance. ... and similar problems are encountered in many other sign languages as well." (Schwager and Zeshan, 2008, p. 513).

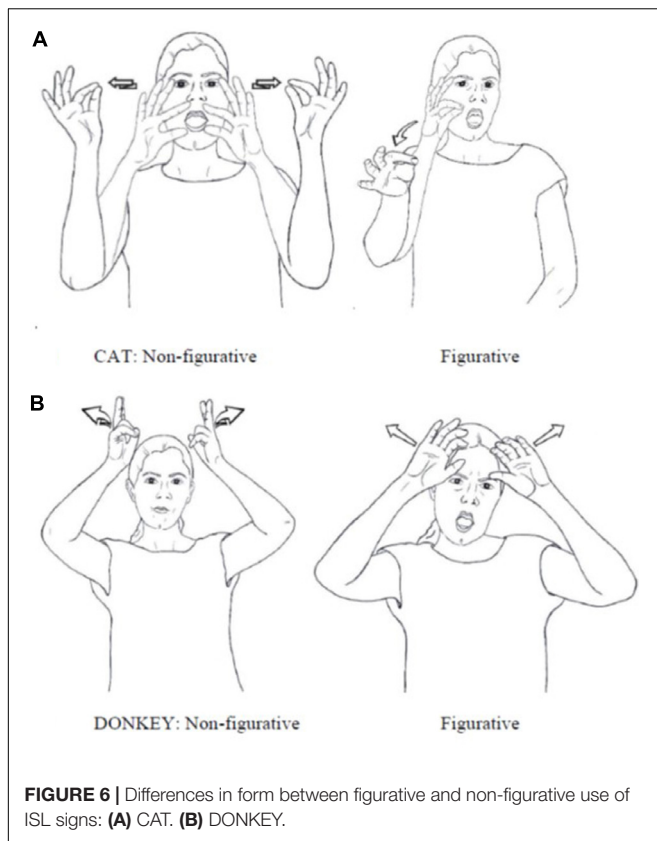
In ISL, for example, a sign such as LONELY may function as an adjective in (15) and as a noun in (16), and this is characteristics of many signs. In many cases, lexical category is assigned according to the function of a sign in a specific syntactic environment rather than as a lexical property of that sign.

- (15) I LONELY 'I am lonely.'
 (16) LONELY MORE MORE WIDE-SPREAD 'Loneliness becomes more and more wide-spread.'

We propose that in the case of nominal metaphor, the noun is reinterpreted as an adjective or adverb, and then it is more readily be construed as dependent. For example, in (12') SNAKE is interpreted as an adjective:

- (12') HE SNAKE 'He (is) snaky/snakelike.'

One piece of evidence for categorical reinterpretation of figurative signs comes from the fact that many metaphorical signs in ISL and ASL have a slightly different form from their non-metaphorical counterparts. This has been observed for ASL by Klima and Bellugi (1979, p. 299) "Figurative extensions of meaning are preferentially accompanied by minimal changes in movement." In ISL, we find that the difference in the quality of movement (as in the sign CAT, **Figure 6A**) is often accompanied by other phonological differences such as the number of hands (non-figurative CAT is two-handed while figurative CAT is one-handed) and handshape (as in DONKEY, **Figure 6B**). However, the quality of movement is crucial here, since it is often associated with differentiating parts of speech in sign languages. For example, nouns and verbs in noun-verb pairs in many sign languages are distinguished by length and quality of movement. This observation was first made for ASL by Supalla and Newport (1978), and then found in other sign languages (see Schwager and Zeshan, 2008; Tkachman and Sandler, 2013 for an overview). Furthermore, ASL has means for deriving verbal/adjectival predicates from nouns. Klima and Bellugi (1979, p. 296) describe a systematic change to the movement of ASL nouns, forming predicates with the meaning of 'to act/appear like X', as in 'to act like a baby' from BABY, 'to seem Chinese' from CHINESE and 'pious' from CHURCH. The derived predicates have a fast and tense movement with restrained onset. Similarly,



differences in movement encode an extended use of signs as sentential adverbials, as in ‘suddenly’ or ‘unexpectedly’ from WRONG, ‘unfortunately’ from TROUBLE.

A second piece of evidence demonstrating that nominal metaphors are indeed interpreted as adjectives is the fact that they can be modified by a degree adverb such as ‘very’ (17a). Such modification is also possible with regular adjectives (17b), but impossible when the noun is used literally (17c).

- (17) (a) HE CAT_{figurative} VERY. ‘He is very “catty”/sly.’
 (b) HE STRONG VERY. ‘He is very strong.’
 (c) *HE CAT_{literal} VERY. ‘He is very cat.’

Back to Similes vs. Metaphors

The proposal that metaphorically interpreted-nouns are reinterpreted as adjectives explains why similes are strongly dispreferred in ISL: syntactically, the preposition LIKE cannot precede an adjective or an adverb.

- (18) JOHN (*LIKE) SNAKE_{figurative}

This is to be contrasted with a spoken language like English, where the figuratively interpreted noun does not change its category, and can therefore unproblematically combine with a preposition:

- (11’) John is (like) a snake.

Note that this explanation is essentially syntactic, and is not dependent on any difference in meaning between metaphors and

similes. Even if metaphors are, indeed, elided similes, the scarcity of similes in sign languages would still be explained in the same way, on the basis of the categorial flexibility of these languages.

There is an important lesson here. In an attempt to demonstrate that metaphors and similes differ in their meaning, we tried to find languages that have metaphors but (almost) no similes. We have, indeed, found such a language—ISL; and yet, we found that the facts of this language have nothing to do with any supposed semantic difference between metaphors and similes. Hence, if anything, our findings provide support for the view that metaphors and similes *are* very close in their meaning.

CONCLUSION AND FUTURE WORK

In the preceding sections, we focused on the differences between manual-visual languages and auditory languages in the expression and use of metaphor and simile. A key issue here is the greater ability of sign languages for iconic expressions. While iconicity provides signs with the ability to represent visual aspects of concepts in a vivid and straightforward way, it also constrains those signs from taking additional, metaphorical meanings, if these are not built on the visual imagery profiled by the sign’s iconicity. As we pointed out, iconicity cannot be inhibited, while metaphorical interpretation is built on inhibition. If both forces play tug-of-war on the same meaning components, the metaphorical interpretation is blocked. The use of similes is further constrained by the categorial shift from nouns to adjectives/adverbs. This shift is made possible by the general flexibility of sign languages regarding lexical categories. But it also constrains the use of similes, since a preposition such as LIKE must be followed by a noun, not an adjective/verb.

We would like to conclude by describing two striking difference between the use of metaphors in signed vs. spoken languages, for which we do not yet have an explanation.

The Expression of Metaphor

We have seen in Section “Categorical Reinterpretation of Figurative Signs” that when a sign is interpreted metaphorically, its form changes slightly. This constitutes an interesting difference between languages in the two modalities: in spoken languages, the main expression of metaphor is through the use of a word in a different semantic domain with an accompanying change in meaning, as in *wave* (*an electromagnetic wave, waves of immigrants, feminism wave, wave of excitement*); in sign languages, in contrast, the main expression of metaphor is in creating new signs (Taub, 2001; Roush, 2016). Sign language abound with metaphorical signs, signs built on both iconic and metaphorical mapping, such as the sign LEARN (Figure 2). Crucially, it is not the case that the sign EAT itself is used metaphorically; rather, the form of the sign is changed in a specific manner—movement toward the signer’s temple rather the mouth—and a *new* sign is formed.

In fact, many (if not most) of the signs denoting abstract concepts in a given sign language are built on this double mapping, as illustrated and exemplified in depth by Taub (2001).

Moreover, this is a very productive way for creating new signs, in everyday use and in sign language poetry. In spoken languages, metaphor is often described as a process of making novel use of existing means: existing lexical items are used to refer to novel concepts by means of metaphorical extensions. In sign languages, this description is not accurate: metaphor is usually not making novel use of existing means, but rather the means for creating novel forms. At present, it is not clear to us how to account for this difference, and we leave it as an open question for future research.

Alternatives to Metaphor

Another difference between languages of the two modalities pertains to providing alternatives to metaphors. Metaphors and similes are often used to create a vivid sensory image. But there are other means for achieving this goal. One alternative way to create vivid imagery is through iconic means, which, as we have seen above, cannot be inhibited.

Since iconicity is much more prevalent in sign languages, we expect to find many instances of vivid iconic representations of the desired visual image, instead of metaphors or similes. This is indeed the case. For example, while the salience of body-parts can inhibit certain metaphorical uses, as we discussed in Section “The Body in Metaphors,” it can also be exploited for specific effects, both in everyday use and in poetic signing. A widespread use of body parts in visual languages is signaling the target domain of metaphorical mapping by articulating a sign close to a specific body part. We saw an example with the sign LEARN, where the head signals that the action encoded in the sign is a mental action. Another example is the sign for BOIL, usually articulated in neutral space. However, when signed close to the chest, the metaphorical site for emotions, it means ‘inner boiling,’ that is, VERY-ANGRY (Figure 7). Changing the sign’s location can be also used creatively. One of our consultants signed the sign DEplete on his bicep instead of in neutral space to convey the meaning of ‘to be exhausted, run out of steam’ (Meir and Sandler, 2008, p. 57). Another consultant created a new sign by signing the sign SHINE close to the eyes, to convey the meaning ‘shining eyes.’ The consultant pointed out that this neologism is more vivid in evoking a mental image of shining eyes than using a metaphor such as ‘*Her eyes were shining stars.*’ The deaf Dutch poet Wim Emerik, in his poem ‘Member of Parliament,’ uses the same technique to convey the idea that the politician consumes the reported news as an automatized bodily function. As the poet depicts the politician eating lunch and reading the newspaper, he changes the location of the verb EAT from the mouth to the eyes, indicating that the politician consumes news as he consumes food (Figure 8) (ibid., 56)⁸.

⁸ Another creative use of the signer’s body is personification, where the poet uses his/her body to ‘become’ the entity (person, animal, object) in the poem. This creates a blend of the characteristics of the poet and characteristics of the focal entity, highlighting how those entities would view the world not only as humans but as signing deaf people, who view the world visually and communicate in signs (Sutton-Space, 2012, p. 1007. See Sutton-Spence and Napoli, 2010 for a thorough description of this device).

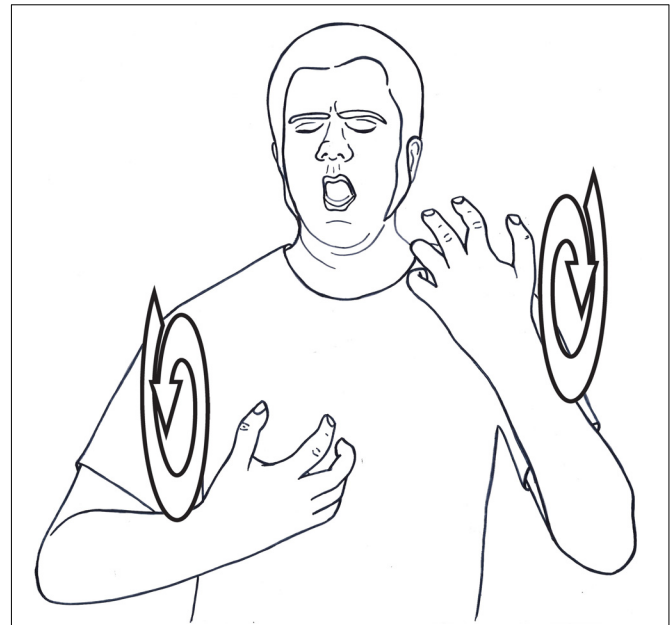


FIGURE 7 | Body parts signal target domain of metaphor: VERY-ANGRY (boiling signed close to chest, signaling the domain of emotions).

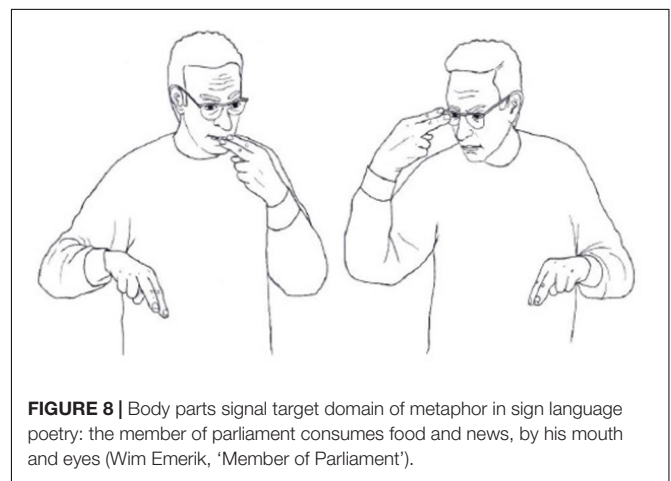


FIGURE 8 | Body parts signal target domain of metaphor in sign language poetry: the member of parliament consumes food and news, by his mouth and eyes (Wim Emerik, ‘Member of Parliament’).

All these examples show how the iconicity of body parts in sign languages can be exploited to create a vivid sensory image, without resorting to explicit or implicit comparisons as in similes and metaphors. Spoken languages cannot exploit body-part terms in the same way, since these terms are not iconic in the spoken modality. Yet iconicity can be used to create sensory image even in the spoken modality. Languages that have a wide array of mimetics, often prefer those to the use of metaphors: “... some types of pain that are mimicked by mimetics in Japanese are expressed by metaphors in other languages (e.g., *gangan* ‘one’s head pounding,’ *kirikiri* ‘one’s stomach splitting,’ *sikusiku* ‘one’s stomach griping’).” (Akita, 2016, p. 155). And Sharlin (2009, p. 5) concludes: “Ideophones, flexible and welcoming to creativity, seem to take the place of other figurative language (simile, metaphor) generally absent in Japanese.” It seems, then,

that creating a vivid sensory image is cherished by language users in general. But different languages have different means for achieving this goal. Affordances of the modality may channel languages to use specific means, e.g., the preference of languages in the signing modality to use iconic expressions. But as the use of mimetics show, it is not only about modality. Even within the same modality, languages may show different preferences. We leave it for future study to investigate the different factors that may lead languages to show preference to one figurative means over another.

A Final Word

In languages in both modalities, metaphor is prevalent, and many metaphorical mappings are shared by both, providing strong support for a universalist view of metaphor and metaphorical mapping in language. But each modality has its own constraints and affordances, shaping the use of metaphors in ways particular to the modality. A cross-modal and cross-linguistic comparison enables us to grasp the central role of metaphors in human expression on the one hand, and the different means that languages provide for carrying out this function.

REFERENCES

- Akita, K. (2016). "A multimedia encyclopedia of Japanese mimetics: a frame-semantic approach to L2 sound-symbolic words," in *Cognitive-Functional Approaches to the Study of Japanese as a Second Language*, eds K. Kabata and K. Toratani (Boston, MA: Walter de Gruyter Inc), 139–168.
- Aronoff, M., Meir, I., and Sandler, W. (2005). The paradox of sign language morphology. *Language* 81, 301–344. doi: 10.1353/lan.2005.0043
- Baus, C., Carreiras, M., and Emmorey, K. (2013). When does iconicity in sign language matter? *Lang. Cogn. Process.* 28, 261–271. doi: 10.1080/01690965.2011.620374
- Bergmann, M. (1979). Metaphor and formal semantic theory. *Poetics* 8, 213–230. doi: 10.1016/0304-422X(79)90021-4
- Bowlde, B. F., and Gentner, D. (2005). The career of metaphor. *Psychol. Rev.* 112, 193–216. doi: 10.1037/0033-295X.112.1.193
- Brennan, M. (1990). *Word Formation in British Sign Language*. Stockholm: Stockholm University Press.
- Chiappe, D. L., and Kennedy, J. M. (2001). Literal bases for metaphor and simile. *Metaphor Symb.* 16, 249–276. doi: 10.1080/10926488.2001.9678897
- Cienki, A., and Müller, C. (2008). "Metaphor, gesture and thought," in *The Cambridge Handbook of Metaphor and Thought*, ed. R. W. Gibbs (Cambridge: Cambridge University Press), 483–501. doi: 10.1017/CBO9780511816802.029
- Cohen, A., and Meir, I. (2014). "Figurative interpretation of dependent and autonomous elements: the view from sign language," in *Proceedings of the High Desert Linguistic Conference*, Albuquerque, NM.
- Cohen, A., and Meir, I. (2015). "Function variables in metaphoric interpretation: evidence from iconicity in sign languages," in *Proceedings of the IATL 31, MIT Working Papers in Linguistics #82E*, ed. A. Bar-Asher Siegal (Haifa: University of Haifa).
- Crasborn, O. (2012). "Phonetics," in *Sign Language: An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll (Boston: De Gruyter), 4–20.
- Croft, W. (2003). "The role of domains in the interpretation of metaphors and metonymies," in *Metaphor and Metonymy in Comparison and Contrast*, eds R. Dirven and R. Pörrings (New York, NY: Mouton de Gruyter), 161–206.
- Emmorey, K. (2014). Iconicity as structure mapping. *Philos. Trans. R. Soc. B* 369:20130301. doi: 10.1098/rstb.2013.0301

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

IM had tragically passed away before the paper was published. She was a brilliant linguist and a wonderful colleague, and will be sorely missed by all who knew her.

FUNDING

This work was partly supported by a grant from the Israel Science Foundation No. 553/04 to IM.

ACKNOWLEDGMENTS

We would like to thank Philip Schlenker and Jeremy Kuhn for insightful discussions, and Debbie Menashe, Sara Lanesman, Yifat Ziv Ben-Zeev and Meir Etdegi, for providing the ISL data on which this study is based, and for helpful discussions of these data. Wendy Sandler is warmly thanked for excellent suggestions and invaluable help.

- Fernandez, P. R. (2007). Suppression in metaphor interpretation: differences between meaning selection and meaning construction. *J. Semant.* 24, 345–371. doi: 10.1093/jos/ffm006
- Forceville, C. (1996). *Pictorial Metaphor in Advertising*. London: Routledge. doi: 10.4324/9780203272305
- Forceville, C. (2005). Addressing an audience: time, place, and genre in Peter Van Straaten's calendar cartoons. *Humor* 18, 247–278. doi: 10.1515/humr.2005.18.3.247
- Forceville, C. (2008). "Metaphor in pictures and multimodal representations," in *The Cambridge Handbook of Metaphor and Thought*, ed. R. W. Gibbs (Cambridge: Cambridge University Press), 462–481. doi: 10.1017/CBO9780511816802.028
- Gentner, D., and France, I. M. (1988). "The verb mutability effect: studies of the combinatorial semantics of nouns and verbs," in *Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence*, eds S. L. Small, G. W. Cottrell, and M. K. Tanenhaus (San Mateo, CA: Morgan Kaufmann Publishers), 343–382.
- Glucksberg, S. (2001). *Understanding Figurative Language: From Metaphors to Idioms*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/978019511095.001.0001
- Glucksberg, S. (2008). "How metaphors create categories — quickly," in *The Cambridge Handbook of Metaphor and Thought*, ed. R. W. Gibbs (Cambridge: Cambridge University Press), 67–83. doi: 10.1017/CBO9780511816802.006
- Glucksberg, S. (2011). "Understanding metaphors: the paradox of unlike things compared," in *Affective Computing and Sentiment Analysis: Emotion, Metaphor and Terminology*, ed. K. Ahmad (New York, NY: Springer), 1–12.
- Glucksberg, S., and Haught, C. (2006a). Can Florida become like the next Florida? When metaphoric comparisons fail. *Psychol. Sci.* 17, 935–938.
- Glucksberg, S., and Haught, C. (2006b). On the relation between metaphor and simile: when comparison fails. *Mind Lang.* 21, 360–378. doi: 10.1111/j.1468-0017.2006.00282.x
- Glucksberg, S., Newsome, M. R., and Goldvarg, Y. (2001). Inhibition of the literal: filtering metaphor-irrelevant information during metaphor comprehension. *Metaphor Symb.* 16, 277–298. doi: 10.1080/10926488.2001.9678898
- Huang, S. (1994). "Chinese as a metonymic language," in *In Honor of William S. Y. Wang: Interdisciplinary Studies on Language and Language Change*, eds M. Chen and O. J. L. Tzeng (Taipei: Pyramid Press), 223–252.
- Johnston, T., and Schembri, A. (1999). On defining lexeme in a signed language. *Sign Lang. Linguist.* 2, 115–185. doi: 10.1075/sll.2.2.03joh

- Jones, L. L., and Estes, Z. (2006). Roosters, robins, and alarm clocks: aptness and conventionality in metaphor comprehension. *J. Mem. Lang.* 55, 18–32. doi: 10.1016/j.jml.2006.02.004
- Kennedy, J. M. (1982). Metaphor in pictures. *Perception* 11, 589–605. doi: 10.1068/p110589
- Kennedy, J. M. (2008). “Metaphor and art,” in *The Cambridge Handbook of Metaphor and Thought*, ed. R. W. Gibbs (Cambridge: Cambridge University Press), 447–461. doi: 10.1017/CBO9780511816802.027
- Klima, E. S., and Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.
- Lakoff, G. (1990). The invariance hypothesis: is abstract reason based on image schemas? *Cogn. Linguist.* 1, 39–74. doi: 10.1515/cogl.1990.1.1.39
- Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar*, Vol. I. Stanford, CA: Stanford University Press.
- Langacker, R. W. (1989). “Absolute construal,” in *Worlds Behind Words: Essays in Honour of Prof. Dr. F. G. Droste on the Occasion of His Sixtieth Birthday*, eds F. J. Hevyaert and F. Steurs (Leuven: Leuven University Press), 65–75.
- Langacker, R. W. (1991). *Foundations of Cognitive Grammar*, Vol. II. Stanford, CA: Stanford University Press.
- Langacker, R. W. (2002). *Concept, Image and Symbol: The Cognitive Basis of Grammar*. New York, NY: Mouton de Gruyter.
- Langdon, R., Davies, M., and Coltheart, M. (2002). Understanding minds and understanding communicated meanings in schizophrenia. *Mind Lang.* 17, 68–104. doi: 10.1111/1468-0017.00190
- Lepic, R., Börstell, C., Belsitzman, G., and Sandler, W. (2016). Taking meaning in hand: iconic motivations in two-handed signs. *Sign Lang. Linguist.* 1, 37–81. doi: 10.1075/sll.19.1.02lep
- Liddell, S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511615054
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.
- Meier, R. P., Comier, K., and Quinto-Pozos, D. (eds) (2002). *Modality and Structure in Signed and Spoken Languages*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486777
- Meir, I. (2010). Iconicity and metaphor: constraints on metaphorical use of iconic forms. *Language* 86, 865–896. doi: 10.1353/lan.2010.0044
- Meir, I. (2012). “Word classes and word formation,” in *Handbook on Sign Language Linguistics*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton De Gruyter), 77–112.
- Meir, I., and Sandler, W. (2008). *Language in Space: A Window onto Israeli Sign Language*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Meir, I., Sandler, W., Padden, C., and Aronoff, M. (2012). *A Dictionary of Al-Sayyid Bedouin Sign Language*. San Diego, CA: University of Haifa and University of California.
- Miller, G. A. (1993). “Images and models, similes and metaphors,” in *Metaphor and Thought*, 2nd Edn, ed. A. Ortony (Cambridge: Cambridge University Press), 357–400. doi: 10.1017/CBO9781139173865.019
- Müller, C., and Cienki, A. (2009). “Words, gestures, and beyond: forms of multimodal metaphor in the use of spoken language,” in *Multimodal Metaphor*, eds C. J. Forceville and E. Urios-Aparisi (Berlin: Mouton de Gruyter), 297–328.
- Nunez, R. (2008). “Conceptual metaphor, human cognition, and the nature of mathematics,” in *The Cambridge Handbook of Metaphor and Thought*, ed. R. W. Gibbs (Cambridge: Cambridge University Press), 339–362. doi: 10.1017/CBO9780511816802.021
- Roush, D. (2016). The expression of the location event-structure metaphor in American sign language. *Sign Lang. Stud.* 16, 389–432. doi: 10.1353/sls.2016.0008
- Sandler, W. (1993). Hand in hand: the roles of the nondominant hand in sign language phonology. *Linguist. Rev.* 10, 337–390. doi: 10.1515/tlir.1993.10.4.337
- Sandler, W. (2006). “Phonology, phonetics, and the nondominant hand,” in *Papers in Laboratory Phonology: Varieties of Phonological Competence*, eds L. Goldstein, D. H. Whalen, and C. Best (Berlin: Mouton-de Gruyter), 185–212.
- Sandler, W. (2013). Dedicated gestures and the emergence of sign language. *Gesture* 12, 265–307. doi: 10.1075/gest.12.3.01san
- Sandler, W., and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139163910
- Schwager, W., and Zeshan, U. (2008). Word classes in sign languages: criteria and classification. *Stud. Lang.* 32, 509–545. doi: 10.1016/j.jbi.2013.06.001
- Sharlin, N. (2009). *Sounds Like: Understanding Japanese Sound Symbolism*. Master’s thesis, Bryn Mawr College, Bryn Mawr, PA.
- Sullivan, K. (2007). *Grammar in Metaphor: A Construction Grammar Account of Metaphoric Language*. Ph.D. dissertation, University of California, Berkeley, CA.
- Sullivan, K. (2009). “Grammatical constructions in metaphoric language,” in *Cognitive Corpus Linguistics*, eds B. Lewandowska-Tomaszczyk and K. Dziwirek (Frankfurt: Peter Lang Publishers), 57–80.
- Supalla, T., and Newport, E. (1978). “How many seats in a chair? the derivation of nouns and verbs in American Sign Language,” in *Understanding Language through Sign Language Research*, ed. P. Siple (New York, NY: Academic Press), 91–132.
- Sutton-Space, R. (2012). “Poetry,” in *Handbook on Sign Language Linguistics*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton De Gruyter), 998–1022.
- Sutton-Spence, R., and Napoli, D. J. (2010). Anthropomorphism in sign languages: a look at poetry and storytelling with a focus on british sign language. *Sign Lang. Stud.* 10, 442–475. doi: 10.1353/sls.0.0055
- Sutton-Spence, R., and Woll, B. (1999). *The Linguistics of British Sign Language: An Introduction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139167048
- Taub, S. F. (2001). *Language from the Body: Iconicity and Metaphor in American Sign Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511509629
- Thompson, R. L., Vinson, D. P., and Vigliocco, G. (2010). The link between form and meaning in British Sign Language: effect of iconicity for phonological decisions. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 1017–1027. doi: 10.1037/a0019339
- Tkachman, O., and Sandler, W. (2013). The noun-verb distinction in two young sign languages. *Gesture* 13, 287–308.
- Utsumi, A. (2007). Interpretive diversity explains metaphor-simile distinction. *Metaphor Symb.* 22, 291–312. doi: 10.1080/10926480701528071
- Utsumi, A. (2011). Computational exploration of metaphor comprehension processes using a semantic space model. *Cogn. Sci.* 35, 251–296. doi: 10.1111/j.1551-6709.2010.01144.x
- van Genabith, J. (2001). Metaphors, logic and type theory. *Metaphor Symb.* 16, 43–57. doi: 10.1080/10926488.2001.9678885
- Vermeerbergen, M., Leeson, L., and Crasborn, O. (2007). *Simultaneity in Sign Languages: Form and Function*. Amsterdam: Benjamins. doi: 10.1075/cilt.281
- Wilcox, P. (2000). *Metaphor in American Sign Language*. Washington, DC: Gallaudet University Press.
- Zbikowski, L. M. (2008). “Metaphor and music,” in *The Cambridge Handbook of Metaphor and Thought*, ed. R. W. Gibbs (Cambridge: Cambridge University Press), 502–524. doi: 10.1017/CBO9780511816802.030

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

IM has been collaborating on research projects with host editors Wendy Sandler and Carol Padden over the past two decades.

Copyright © 2018 Meir and Cohen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Visual Iconicity Across Sign Languages: Large-Scale Automated Video Analysis of Iconic Articulators and Locations

Robert Östling^{1*}, Carl Börstell^{1,2} and Servane Courtaux³

¹ Department of Linguistics, Stockholm University, Stockholm, Sweden, ² Centre for Language Studies, Radboud University, Nijmegen, Netherlands, ³ École Nationale Supérieure de Techniques Avancées, ParisTech, Paris, France

OPEN ACCESS

Edited by:

Marianne Gullberg,
Lund University, Sweden

Reviewed by:

Amber Joy Martin,
Hunter College (CUNY), United States
Marcus Perlman,
University of Birmingham, United
Kingdom

*Correspondence:

Robert Östling
robert@ling.su.se

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 15 November 2017

Accepted: 25 April 2018

Published: 15 May 2018

Citation:

Östling R, Börstell C and Courtaux S
(2018) Visual Iconicity Across Sign
Languages: Large-Scale Automated
Video Analysis of Iconic Articulators
and Locations. *Front. Psychol.* 9:725.
doi: 10.3389/fpsyg.2018.00725

We use automatic processing of 120,000 sign videos in 31 different sign languages to show a cross-linguistic pattern for two types of iconic form–meaning relationships in the visual modality. First, we demonstrate that the degree of inherent plurality of concepts, based on individual ratings by non-signers, strongly correlates with the number of hands used in the sign forms encoding the same concepts across sign languages. Second, we show that certain concepts are iconically articulated around specific parts of the body, as predicted by the associational intuitions by non-signers. The implications of our results are both theoretical and methodological. With regard to theoretical implications, we corroborate previous research by demonstrating and quantifying, using a much larger material than previously available, the iconic nature of languages in the visual modality. As for the methodological implications, we show how automatic methods are, in fact, useful for performing large-scale analysis of sign language data, to a high level of accuracy, as indicated by our manual error analysis.

Keywords: iconicity, sign language, location, two-handed signs, semantics, lexical plurality, automated video processing, typology

1. INTRODUCTION

1.1. Iconicity Across Languages and Modalities

The traditional view of the linguistic lexical unit has been that it is created by combining meaningless units (phonemes) into a meaning-bearing form (morpheme/word) which is semantically not compositional or even motivated, i.e., it is arbitrary (de Saussure, 1916). Arbitrariness of word forms has even been used as a criterion for what constitutes *language* (Hockett, 1960). However, the view of the building blocks of language as being entirely arbitrary has later been questioned after it has been found—across languages—that both units smaller than words and words themselves may exhibit non-arbitrariness. The clearest case of non-arbitrariness is *iconicity*, the direct form–meaning association in which the linguistic sign resembles the denoted referent in form, which has been found across languages in both the spoken and the signed modality (Perniss et al., 2010; Meir and Tkachman, 2018). However, it is often claimed that signed languages are more iconic than spoken languages, because the former are—due to the visual modality—“richer in iconic devices” (Meir and Tkachman, 2018).

For spoken languages, sensory imagery appears to be particularly associated with various forms of iconic sound symbolism (e.g., *splash*, *beep*) (e.g., Dingemanse, 2012; Schmidtkne et al., 2014; Kwon and Round, 2015; Dingemanse et al., 2016; Winter et al., 2017), and has also been recreated in experimental settings (e.g., Köhler, 1929; Lockwood and Dingemanse, 2015; Cuskley et al., 2017; Fort et al., 2018). Urban (2011) and Blasi et al. (2016) demonstrate that there are a number of iconic form–meaning mappings across large samples of the world’s spoken languages, such as the mapping between phonemes and its associated articulatory body part (e.g., /n/ with “nose,” /l/ with “tongue,” and bilabials with “lip”), or the mapping between phonemes and physical properties (e.g., /i/ with “small”). Although iconic expression is often concrete, it may also be extended to abstract senses (Auracher, 2017). Thus, iconicity seems to be an integral part of language (Dingemanse et al., 2015), which has been found to be a facilitating element when acquiring a language, regardless of the modality (e.g., Thompson et al., 2012; Monaghan et al., 2014; Lockwood et al., 2016; Ortega, 2017). In their overview of research on (non-)arbitrariness in the vocabularies of human languages, Dingemanse et al. (2015) conclude that there are trade-offs in making use of arbitrariness and non-arbitrariness (e.g., iconicity). Whereas arbitrariness offers fewer constraints on form, it makes it harder for users to learn; and whereas iconicity facilitates learning, it may restrict the forms and abstraction possibilities of the language. We hope that our present work in the visual modality can contribute toward determining the parameters involved in this trade-off.

The type of iconicity that is possible is partly dependent on the modality. For example, whereas quantity (size, plurality, intensity) is easy to depict iconically with either spoken or signed forms (e.g., by duration or reduplication), spoken language is better for depicting sound, and signed language is better for depicting visual properties and space (Dingemanse et al., 2015, 608)¹. In this paper, we are mainly interested in the visual iconicity found among signed languages.

1.2. Sign Language Iconicity

Already some of the earliest research on signed languages acknowledged the iconic motivation found in many signs (Klima and Bellugi, 1979). That is, the form of a sign is directly motivated by visual properties of its referent. This is a quite uncontroversial claim today, and there is a growing body of work looking at the interaction between iconicity and the structure of signed language (e.g., Taub, 2001; Meir, 2010; Lepic, 2015). However, it has been argued that there is a language-dependent factor in the iconicity of signs, such that signers tend to rate the signs of their own language as more iconic than the signs with corresponding meanings in another sign language (Occhino et al., 2017). Those looking at iconicity as a factor shaping the very structure of signed languages have noted that individual form features of a lexical sign can provide separate parts of the combined semantics of the whole. For instance, the handshape

may be used to describe size and shape properties of an entity, by letting the hand represent either the entity itself, or the hand as it handles the entity (Padden et al., 2013). Similarly, the type of movement in the articulation of a sign may be motivated, such as having movement manner, duration, and onset/offset encode lexical aspect iconically, such as distinct end movements being associated with telicity (Grose et al., 2007). In a model developed by Taub (2001), the form and meaning of signs can be formalized as a so-called *double mapping* in which form parameters are mapped to the concrete source of a metaphor, which in turn are mapped onto a metaphorical target. Using this model, Meir (2010) shows how the Israeli Sign Language sign for “learning” makes use of the metaphor UNDERSTANDING IS GRASPING. In this sign, the handshape represents holding an object, which is interpreted metaphorically as considering an idea, and the sign location (forehead) represents the head, which is metaphorically linked to the location of the mind².

In this paper, we are specifically interested in two broad phonological parameters of signs: number of hands and sign location. Both of these parameters have been found to contribute to the iconicity of signs.

Any sign can be produced with either *one or two hands*. This dichotomy has initially mostly been treated merely as a phonological feature of signs (van der Hulst, 1996), perhaps because the distribution of one- and two-handed signs across sign language lexicons seems to be 50/50 (Börstell et al., 2016; Crasborn and Sáfár, 2016), suggesting a random distribution³. However, it has been shown that the number of hands used in a sign can be attributed to meaning, based on the iconic mapping between the articulators (e.g., the hands) and (parts of) a referent, for example as plural/reciprocal alternations (Pfau and Steinbach, 2003, 2006, 2016) or multiple entities (e.g., Dudis, 2004; Zwitserlood et al., 2012). However, the use of two hands for plural expression has also been shown to work on the lexical level. For example, Lepic et al. (2016) showed that while the distribution between one- and two-handed signs in any random sign language lexicon appears balanced and arbitrary, there is significant overlap in which meanings are encoded by two-handed signs *across* languages. Using what the authors term *articulatory plurality*, Börstell et al. (2016) argue that sign languages are able to map plural referents onto the plural articulators. Börstell et al. (2016) and Lepic et al. (2016) show that sign languages favor two-handed sign forms across languages to represent lexically plural concepts. Lexical plurals are concepts that carry some form of inherent plurality in their semantics, for example *reciprocals*, events/relationships necessarily involving multiple participants (“kiss,” “argue,” “friend”) (Haspelmath, 2007; Acquaviva, 2008), and *mass/dual/plural groups or objects*, involving multiple members/parts (“army,” “socks,” “gloves”) (Koptjevskaja-Tamm, 2004; Acquaviva, 2008, 2016; Wisniewski,

²In her paper, other form parameters (e.g., movement) of the sign are also represented as having a double mapping with iconicity and metaphor (Meir, 2010, p. 877).

³One-handed signs seem to constitute a higher proportion of sign *tokens* in corpus data, thus the 50/50 distribution is mainly relevant for sign *types*, i.e., in a dictionary (Crasborn and Sáfár, 2016, p. 244).

¹Note that visual iconicity is present also in the multimodal communication of speakers of spoken languages, as many co-speech gestures are known to be iconic (Poggi, 2008).

2010; Lauwers and Lammert, 2016; Mihatsch, 2016). The association between plurality and two-handed forms is argued to be an iconic mapping in the same domain as the association between repeated or longer word forms and quantity/plurality (Dingemanse et al., 2015), based on the metaphor MORE OF FORM IS MORE OF CONTENT (Lakoff and Johnson, 1980, p. 127).

Another parameter of the lexical sign is its *location* (or, *place of articulation*), which means the place in signing space, on or around the signer's own body, at which a sign is produced. A sign location may be lexically specified or modified. For example, any lexical sign has a lexically specified location—e.g., the signs EAT and SAY are typically signed at the mouth, in an iconic fashion (Frishberg and Gough, 2000; Taub, 2001). However, the location may sometimes be altered in order to indicate associations with referents localized in signing space (Cormier et al., 2015; Occhino and Wilcox, 2016). Since the lexical location may be iconic (e.g., EAT at the mouth), there may be restrictions to the possible modifications of locations in a sign. For example, if the location is iconic, modifying the location results in a loss of (metaphorical) iconic mapping which may be disallowed by the language (Meir, 2010; Meir et al., 2013). In this paper, however, we focus on the issue of location only as a lexically specified parameter of a sign, in order to evaluate quantitatively, within and across sign languages, to what extent this parameter is iconically motivated. Location has long been known to constitute a possible iconic parameter of lexical signs. Locations may be directly or metaphorically associated with a certain meaning—i.e., locations may be iconic. For instance, the forehead is associated with cognition, whereas the chest is associated with emotion (Brennan, 1990, 2005). Although this is a well-known property of lexical locations, it has never been quantified to any larger extent. An attempt at quantifying the form–meaning mapping of sign locations was made by Börstell and Östling (2017) by linking the manually annotated Swedish Sign Language dictionary to a semantic dictionary. This showed that signs of certain semantic domains (e.g., “think,” “see,” “eat”) were more prominent in certain locations (forehead, eyes, and mouth/belly, respectively) than signs in general.

In this paper, we aim to investigate the *number of hands* and *location* in signs and their association with specific semantics. We do this using automated methods on a large dataset with 120,000 videos from a sample of 31 different sign languages, using the parallel sign language dictionary *Spread the Sign* 2012⁴. We specifically hypothesize that (1) the iconic mapping strategy of articulatory plurality, specifically using two-handed sign forms to represent plural meanings, is employed across languages, and (2) sensory and body part-related meanings will be iconically articulated at their associated locations on the body across languages.

2. MATERIALS AND METHODS

We use data from the online parallel dictionary *Spread the Sign* 2012, which contains a total of 31 sign languages with sign entries,

⁴The *Spread the Sign* dictionary has recently been used to create an online database of certain iconic patterns across a sign languages (Kimmelman et al., 2018).

TABLE 1 | The 31 sign languages in the dataset.

American Sign Language	Estonian Sign Language	Latvian Sign Language
Austrian Sign Language	Finnish Sign Language	Lithuanian Sign Language
Belarusian Sign Language	French Sign Language	Polish Sign Language
Brazilian Sign Language	German Sign Language	Portuguese Sign Language
British Sign Language	Greek Sign Language	Romanian Sign Language
Bulgarian Sign Language	Icelandic Sign Language	Russian Sign Language
Chinese Sign Language	Indian Sign Language	Spanish Sign Language
Croatian Sign Language	International Sign	Swedish Sign Language
Cuban Sign Language	Italian Sign Language	Turkish Sign Language
Czech Sign Language	Japanese Sign Language	Ugandan Sign Language
		Ukrainian Sign Language

and roughly 300,000 videos of individual signs. **Table 1** shows the sign languages included in our dataset. The language sample is not typologically balanced, showing a clear bias toward European or European-derived sign languages. The issue of genealogical relatedness and contact is notoriously difficult when it comes to sign languages. There is little research on the topic, and most classifications are based on either historical sources of contact (usually concerning deaf education) or lexicostatistical comparisons classifying languages based on lexical similarity (cf. Brentari, 2010; Jepsen et al., 2015). A basic attempt at a classification is found under the *Sign Language* family in the *Glottolog* database, according to which several of the languages in our dataset are categorized as part the same language group (e.g., Swedish Sign Language and Finnish Sign Language, or German Sign Language and Polish Sign Language) (Hammarström et al., 2017). In this study, we are not concerned with the potential relatedness between languages. The aim is to explore iconic properties of any language in the visual modality and the possible patterns that may be discerned. However, it should be noted that some patterns of metaphorical iconicity found here may be influenced by having a European/Western-biased language sample.

2.1. Aims

In this study, we use computer vision techniques to infer the main areas of hand activity for individual sign videos. This allows us to study visual iconicity in sign languages in several ways: by comparing patterns of articulation and iconicity between different semantic concepts and categories within the same language, or by comparing signs for the same concept or in the same semantic category across different languages in order to see possible patterns across the sign languages of our sample. Our two main research questions are the following:

1. What is the distribution of one- vs. two-handed signs across sign languages?
 - a. Is the general distribution of one- vs. two-handed signs 50/50, when looking at a semantically diverse set of items, as found by Lepic et al. (2016) and Börstell et al. (2016)?
 - b. Is lexical plurality associated with two-handed signs, as argued by Börstell et al. (2016)?

2. Is there a cross-linguistic pattern of sign locations being iconically associated with specific meanings, as found by Börstell and Östling (2017) for Swedish Sign Language (e.g., “think” associated with the forehead)?

We approach these two main research questions in two studies. Study 1 deals with identifying the number of articulators (i.e., one- vs. two-handed signs) and correlating this with plurality. Study 2 deals with the visualization of sign location based on hand activity and the correlation with semantics. Thus, the two studies aim to show the extent to which visual iconicity is found across sign languages with regard to articulators and locations. The studies also provide us with the possibility of evaluating how signed language can be analyzed with the help of automated video processing methods. By utilizing such methods, we are able to quantify some of the claims about sign language iconicity previously investigated with much smaller datasets in terms of the number of languages and the number of signs involved.

2.2. Data Processing

Video files were downloaded from the *Spread the Sign* public website, along with metadata on the language used, as well as the name of the concept and the concept’s category (in English)⁵. As the first step in our processing chain, we used the body pose estimation model of Cao et al. (2016) to identify the position of wrists and elbows for each video frame. We used the model file published by the authors, which has been trained on the COCO dataset (Lin et al., 2014)⁶. While their model can identify the poses of multiple humans in a frame and thus is much more general than needed here, we found that it is highly accurate in identifying the required body parts. This step required about two months of computing time on a single GPU.

Since the body pose estimation model is not directly trained to detect hands, we estimate their location by extrapolating the elbow–wrist line outwards by half the distance between the elbow and wrist. As shown in **Figure 1**, this assumes a hand position based on a straight wrist joint and the lower knuckles (the metacarpophalangeal joints) as the center of the hand, which is a fairly close approximation in most cases, but discards any possible wrist flexion. Furthermore, since there is considerable variation in body shape and camera distance, we normalize the coordinates such that the averaged location of the signer’s nose is at origo, the x axis is scaled by the mean distance between the shoulders, and the y axis is scaled by the mean distance between the nose and the neck (**Figure 2**). This normalization is performed per video.

2.3. Study 1: Number of Articulators

Our main interest is to explain *why* certain signs are two-handed while others are not. In particular, based on previous studies (Börstell et al., 2016; Lepic et al., 2016) we expect that lexical plurality is an important semantic component in predicting whether a sign is two-handed or not. Although lexical plurality has been researched extensively across many languages (e.g.,

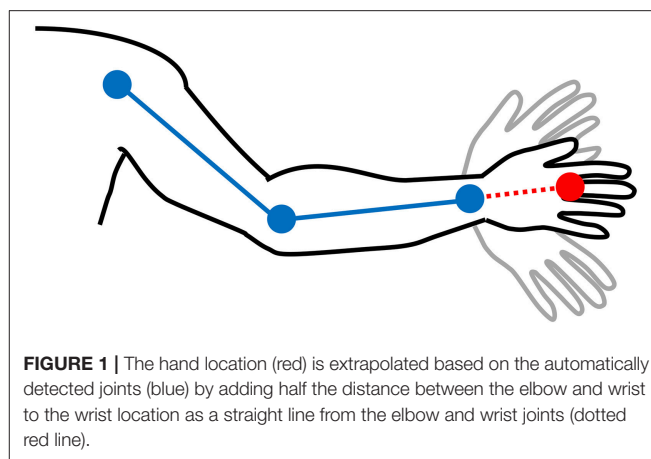


FIGURE 1 | The hand location (red) is extrapolated based on the automatically detected joints (blue) by adding half the distance between the elbow and wrist to the wrist location as a straight line from the elbow and wrist joints (dotted red line).

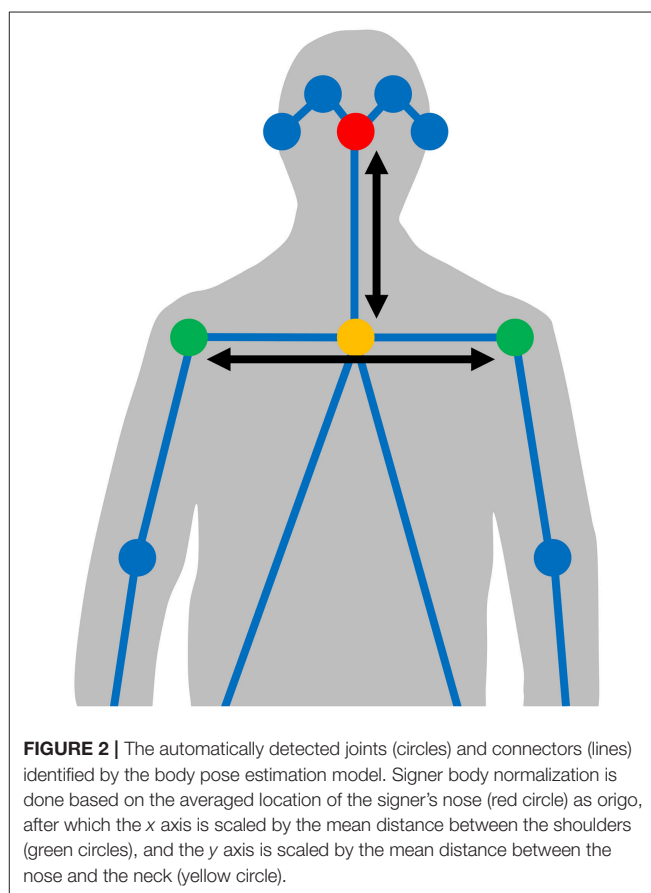


FIGURE 2 | The automatically detected joints (circles) and connectors (lines) identified by the body pose estimation model. Signer body normalization is done based on the averaged location of the signer's nose (red circle) as origo, after which the x axis is scaled by the mean distance between the shoulders (green circles), and the y axis is scaled by the mean distance between the nose and the neck (yellow circle).

Acquaviva, 2008, 2016; Lauwers and Lammert, 2016), we include here data from a plurality rating task, in order to account for plurality—or, specifically the *perceived* plurality—of concepts as a scalar property⁷. For this, we designed a questionnaire as described in section 2.3.2 below. Another possible explanation that we explore is the influence of frequency, since the principle of economy suggests that high-frequency concepts should have

⁵<http://www.spreadthesign.com/>

⁶Common Objects in Context (COCO) is a dataset of 328,000 images of basic objects that have been labeled for developing object recognition models.

⁷We thank one of the reviewers for this suggestion.

forms that are less costly to articulate, which in turn could influence the number of articulators used for specific signs (cf. Crasborn and Sáfár, 2016, p. 244). For this, we used the lemma frequency of the concept's English name in the British National Corpus (Leech et al., 2001).

2.3.1. Data Processing

We estimate the number of articulators by calculating the total length of the path traced by each hand (in the normalized coordinate system) during a sign. If the path length of one hand exceeds the other's by a factor of more than 3, we classify the sign as one-handed. All other signs are classified as two-handed. In rare cases, the video processing stage fails to identify the hand(s). To deal with this we discard all signs with hands detected in <10 frames.

2.3.2. Plurality Rating Questionnaire

Since little quantitative data exists on lexical plurality, we sent out an online questionnaire and collected plurality ratings from respondents ($N = 23$; 10 female, 13 male; mean age 29, SD 10), mainly those without a linguistics background. Respondents were asked to rate the *lexical plurality* of 100 concepts, 50 of which were taken from a previous study (Börstell et al., 2016), collected from various sources (Attarde, 2007; Haspelmath, 2007; Wisniewski, 2010) identifying them as lexically plural concepts, and 50 which were random concepts frequency matched (pairwise) to the lexically plural concepts in another study (Börstell et al., 2016). Lexical plurality is defined in the questionnaire as “whether or not there is some inherent plural meaning of the concept (e.g., involving multiple parts/participants/events).” Concepts were presented in random order, and respondents were asked to provide ratings on a discrete scale from 1 (“not at all plural”) to 7 (“definitely plural”). For our analysis, 19 concepts from the questionnaire were excluded for one of two reasons:

1. 5 concepts: Body parts for which Börstell et al. (2016) used a plural form (e.g., “eyes”) while *Spread the Sign* uses a singular form (e.g., “eye”).
2. 14 concepts: Missing data from *Spread the Sign*.

Thus, a total of 81 concepts with both plurality ratings and number of hands in *Spread the Sign* remained to be used in our analysis.

2.3.3. Statistical Model

We model our data in the following way. The plurality rating of respondent i for concept c , $R_{c,i}$, is assumed to be an independent draw from a normal distribution with standard deviation σ^Q and mean $(1 + 6p_c) + r_i$, that is, $R_{c,i} \sim N(1 + 6p_c + r_i, \sigma^Q)$. Note that the unobserved ‘true’ plurality p_c is a continuous variable on the interval $[0, 1]$ while $R_{c,i} \in \{1, 2, 3, 4, 5, 6, 7\}$. To account for individual bias among respondents, we add an individual noise term $r_i \sim N(0, \sigma^R)$ for respondent i . We assume a logistic regression model for the probability of a sign $S_{c,l}$, expressing concept c in language l , being two-handed: $P(S_{c,l} = 1) = \sigma(\beta_p p_c + \beta_f f_c + \alpha_l^L + \alpha_c^C + a)$, where the logistic function $\sigma(x) = 1/(1 + e^{-x})$. The unobserved variables of the

TABLE 2 | Unobserved variables in our statistical model.

Variable	Prior	Description
β_p	$N(0, 5)$	Plurality coefficient
β_f	$N(0, 5)$	Frequency coefficient
p_c	$U(0, 1)$	Plurality of concept c
r_i	$N(0, \sigma^R)$	Effect of rater i on rating
α_l^L	$N(0, \sigma^L)$	Effect of language l on number of hands
α_c^C	$N(0, \sigma^C)$	Effect of concept c on number of hands
a	$N(0, 5)$	Intercept term
σ^Q	$\text{Exp}(1/5)$	Standard deviation of $R_{c,i}$
σ^R	$\text{Exp}(1/5)$	Standard deviation of α^R
σ^L	$\text{Exp}(1/5)$	Standard deviation of α^L
σ^C	$\text{Exp}(1/5)$	Standard deviation of α^C

TABLE 3 | Observed variables in our statistical model.

Variable	Description
$R_{c,i}$	Plurality rating (integer 1–7) of concept c from respondent i
$S_{c,l}$	Number of hands (1 = two, 0 = one) used in the sign expressing concept c in language l
f_c	log-frequency in British National Corpus of concept c , scaled to $[0, 1]$

model are listed in **Table 2**, and the data are summarized in **Table 3**. Weakly informative priors are used throughout, since we do not have strong prior knowledge to further constrain the model. Logistic regression coefficients are unlikely to have absolute values much above 5, so we use $N(0, 5)$ priors. Standard deviations of either concept/language-specific regression terms or rating scores are also unlikely to be much above 5, so we use exponential priors with $\lambda = 1/5$. For inference, we used the NUTS sampler implemented in the Stan software package (Carpenter et al., 2017) to run four independent chains with 5,000 burn-in iterations followed by 5,000 sampling iterations. All parameters have the Gelman-Rubin statistic $\hat{R} < 1.01$, indicating convergence.

2.4. Study 2: Sign Locations

For visualizing articulator activity over a large number of signs, we also use the normalized coordinates of the hands. These are used to trace the path of each sign individually on a grid. The path is then blurred to reflect the uncertainty in our estimate of the hand location, by convolving it with a Gaussian function ($\sigma = 0.03125x$, where x is the horizontal resolution of the figure), and placed on top of a silhouette positioned as a reference point in relation to the positions of the body pose joint coordinates (see **Figure 2**). The values on the grid are then averaged over the group of signs to be visualized together. In our visualizations we color the activity of the left hand in blue and the right hand in red. The final strength of each color in the visualizations represents the mean amount of time, across languages, each hand spends at a certain location.

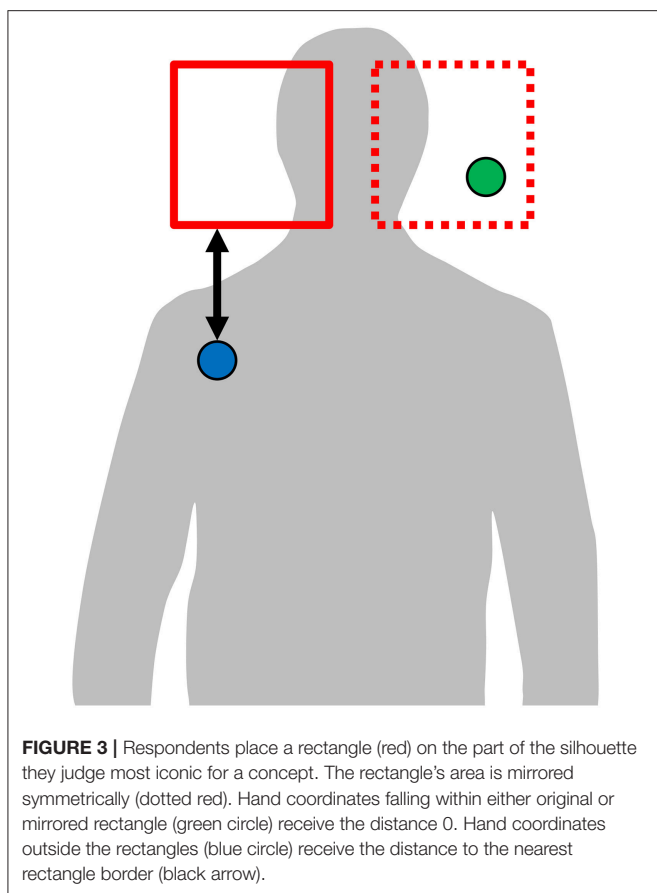


FIGURE 3 | Respondents place a rectangle (red) on the part of the silhouette they judge most iconic for a concept. The rectangle's area is mirrored symmetrically (dotted red). Hand coordinates falling within either original or mirrored rectangle (green circle) receive the distance 0. Hand coordinates outside the rectangles (blue circle) receive the distance to the nearest rectangle border (black arrow).

2.4.1. Quantifying Iconicity Through Location Ratings

To demonstrate that systematicity in sign locations across languages is in fact attributed to iconicity, we define a quantitative measure of iconicity. To obtain data, we created a computer-based visual questionnaire for *location ratings*. Respondents ($N = 10$; 6 female, 4 male; mean age 41, SD 13), none of which reported knowledge of any sign language, were instructed to place a rectangle on a body silhouette (see **Figure 3**). This is the same silhouette we use for visualization of the hand activity across concepts. Respondents were presented, in random order, with the name of the concept in English and the silhouette, and were instructed to place a single rectangle—the size of which could be controlled by the respondent—on the part of the image that they most strongly associate with the concept. To ensure symmetry, the rectangle is mirrored so that any part covered in the left half of the body is also covered on the right half, and vice versa. We refer to the resulting rectangle(s), either one or two, as a *location rating*. The *iconicity score* of a sign with respect to a location rating is then computed in the following way:

1. Guess the dominant hand by choosing the hand with the longest trajectory length during the sign. In case of symmetric two-handed signs, the choice of dominant hand will be arbitrary, but since the computations are symmetric this will not affect the result.

2. All video frames without the dominant hand are removed, and then the first 20% and the last 20% of remaining frames are removed, because they tend to include movements from and to a neutral position.
3. The *iconicity score* is now computed as the negative mean distance, over all remaining video frames, between the dominant hand and the closest rectangle in the location rating. If the hand is inside a rectangle, the distance is zero for that frame. Otherwise, Euclidean distance between the hand and the closest point of any rectangle is used (see **Figure 3**). Thus, the highest possible iconicity score (zero) is obtained if the whole sign is articulated *entirely* inside the space marked by *all* respondents in the location ratings. Lower values (that is, increasingly negative) of the iconicity score indicate increasing divergence from the location ratings.

To compute the iconicity score for a concept, we compute the $N \times M$ iconicity scores for each combination of the N languages that have a sign for the concept and the M location ratings for that concept, and use the mean of these values. Thus, the higher the iconicity score, the closer the articulation of the signs are to the areas indicated by the respondents.

The iconicity score is difficult to interpret out of context. For this reason, we also compare the location ratings to each concept in a vocabulary list, to estimate the level of chance similarity. We use the Swadesh list from Study 1 (see section 2.3), with the concept FOOD added in order to ensure that all the concepts in this study are covered. If iconicity is a significant factor, we expect the location ratings of a particular concept to be closer (have a higher iconicity score) to signs expressing the same concept, than to unrelated signs.

3. RESULTS

3.1. Study 1: Number of Articulators

3.1.1. Processing Quality

To assess the soundness of our method, we decided to manually annotate 20 randomly sampled signs for each language: 10 that were classified as one-handed, and 10 that were classified as two-handed⁸. All the sampled concepts are in the Swadesh list. With this information, we are able to estimate the precision for each category (one-handed and two-handed). While statistical power is limited due to the small sample, we can easily identify two languages for which the automatic processing fails completely: Czech Sign Language and Russian Sign Language. For these languages, only 3 of the 10 signs classified as two-handed are in fact two-handed. For two other languages, British Sign Language and Portuguese Sign Language, 7 of 10 signs classified as two-handed are really two-handed. All other languages contained at most 2 errors per group of 10 for two-handed sign detection, and all languages (including the ones above) have at most 2 errors per group of 10 for one-handed sign detection. Qualitatively, we found that these errors were mainly due to camera setup in these languages, in which the non-articulating hand was partly

⁸Cuban Sign Language was excluded from this manual inspection, since there are <10 signs each in the one- and two-handed classification group, respectively.

TABLE 4 | Number of one-handed (**1H**) and two-handed (**2H**) signs according to the *Spread the Sign* data for concepts in the Swadesh list (**Core vocabulary**) and for all concepts with a single-word English translation (**Extended vocabulary**).

Sign language	Core vocabulary			Extended vocabulary		
	1H	2H	2H/(1H+2H) (%)	1H	2H	2H/(1H+2H) (%)
American Sign Language	102	130	56.0	2,216	3,535	61.5
Austrian Sign Language	100	133	57.1	1,527	3,942	72.1
Belarusian Sign Language	56	89	61.4	738	3,003	80.3
Brazilian Sign Language	126	83	39.7	1,302	1,715	56.8
<i>*British Sign Language</i>	64	170	72.6	715	4,829	87.1
Bulgarian Sign Language	85	67	44.1	588	841	58.9
Chinese Sign Language	83	93	52.8	876	2,977	77.3
Croatian Sign Language	30	39	56.5	402	919	69.6
Cuban Sign Language	6	9	60.0	22	34	60.7
<i>*Czech Sign Language</i>	23	210	90.1	248	5,260	95.5
Estonian Sign Language	100	136	57.6	1,468	4,097	73.6
Finnish Sign Language	17	12	41.4	121	92	43.2
French Sign Language	82	139	62.9	1,184	3,833	76.4
German Sign Language	109	122	52.8	1,815	3,711	67.2
Greek Sign Language	12	25	67.6	483	950	66.3
Icelandic Sign Language	118	114	49.1	1,847	3,754	67.0
Indian Sign Language	91	98	51.9	626	1,714	73.2
International Sign	39	50	56.2	185	291	61.1
Italian Sign Language	102	132	56.4	1,136	4,450	79.7
Japanese Sign Language	79	116	59.5	660	2,033	75.5
Latvian Sign Language	100	134	57.3	1,720	3,905	69.4
Lithuanian Sign Language	117	113	49.1	1,921	3,572	65.0
Polish Sign Language	103	120	53.8	1,545	3,799	71.1
<i>*Portuguese Sign Language</i>	106	117	52.5	1,527	3,669	70.6
Romanian Sign Language	96	91	48.7	1,943	1,941	50.0
<i>*Russian Sign Language</i>	71	156	68.7	1,446	3,980	73.4
Spanish Sign Language	121	110	47.6	1,670	3,759	69.2
Swedish Sign Language	109	115	51.3	1,924	3,582	65.1
Turkish Sign Language	113	117	50.9	1,976	3,412	63.3
Ugandan Sign Language	15	19	55.9	75	135	64.3
Ukrainian Sign Language	118	115	49.4	1,739	3,556	67.2
Total	2,493	3,174	56.0	35,645	87,290	71.0

The proportion of two-handed signs [$2H/(1H+2H)$] is also given. Languages marked with asterisk have systematic processing errors.

in frame (and moving slightly) during one-handed signs, thereby confusing the body pose estimation model. Overall precision for one-handed sign detection is 95.0% (13 errors in 260 signs), and 95.8% (11 errors in 260 signs) for two-handed sign detection if the four problematic cases mentioned above are removed. Without removing these, precision for two-handed sign detection drops to 89.7% (31 errors in 300 signs) while one-handed sign detection remains high at 95.7% (13 errors in 300 signs). We take this as a proof of validity for our automated method in identifying the number of hands in sign videos.

3.1.2. Distribution of One- vs. Two-Handed Signs

Table 4 shows the distribution of one- and two-handed signs in each language, for both *core vocabulary* and *extended vocabulary*. We define the *core vocabulary* to include all signs available in the

data that represent concepts in the Swadesh list from Lepic et al. (2016), in total 5,667 signs for 195 concepts. Going beyond the core vocabulary is problematic, since two-handedness becomes a less meaningful property when dealing with compound signs or whole phrases, and *Spread the Sign* does not contain enough information to distinguish these from simple signs. In order to obtain an *extended vocabulary* of mostly non-compound signs, we use all signs for concepts with single-word translations in an isolating language (English). This results in 122,935 signs, an amount that would be very time-consuming to classify manually, hence our automatic processing demonstrates its usefulness.

From these results, we see that the proportion of two-handed signs in the extended vocabulary (71.0%) is much higher than the corresponding figure for the core vocabulary (56.0%, cf.

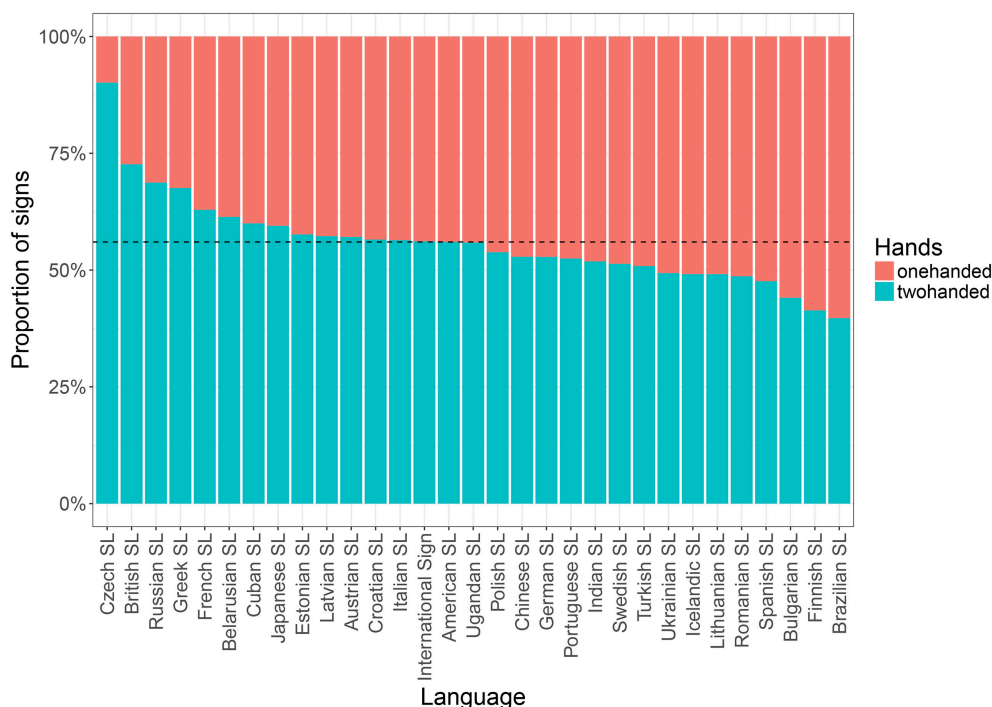


FIGURE 4 | Distribution of one- vs. two-handed signs ($n = 5,667$) across languages for concepts ($n = 195$) in the Swadesh list; dotted line showing the mean proportion of two-handed signs across all signs (56.0%). Note that all languages in our data are included here, i.e., also those excluded from our analysis due to unreliable video processing (Czech SL, British SL, Russian SL, and Portuguese SL).

Figure 4), although still lower than in the list of lexical plurals (81.8%, cf. Figure 5)⁹. This may be explained by lexical frequency and articulatory economy. For instance, Crasborn and Sáfár (2016) show that the while the distribution of one- vs. two-handed signs in Sign Language of the Netherlands is more or less balanced in lexical databases, there is a bias toward one-handed signs in corpus tokens. The authors suggest that ease of articulation may cause lexically two-handed signs to be produced as one-handed signs (Crasborn and Sáfár, 2016, p. 244). In fact, it may even be possible that frequency effects on phonetic reduction pushes toward one-handed articulation with frequent signs, similarly to how frequent signs have been shown to be the most reduced in terms of sign duration (Börstell et al., 2016). However, seeing as other sign databases of similar size to *Spread the Sign* (per individual language) have been shown previously to exhibit a near 50/50 distribution of one- vs. two-handed signs (Börstell et al., 2016, p. 393), it is also possible that the selection of concepts in the *Spread the Sign* project affects the distribution. As stated on their website, one objective of the project was to facilitate vocational training exchanges between countries, and possibly the inclusion of vocational school terminology results in a higher proportion of complex concepts that require compound, phrasal, and/or depicting constructions, which are more likely to be encoded (in part) by two-handed forms.

⁹If the four languages with systematic errors are removed, the mean proportion of two-handed signs drops to 53.1%—i.e., even closer to an even distribution.

3.1.3. The Influence of Plurality

Based on previous research (Börstell et al., 2016) and the results obtained here (see Table 4 and Figure 4), we expect the—at least core—vocabulary of any sign language to exhibit a close to 50/50 distribution between one- and two-handed signs. As shown by Börstell et al. (2016) for a sample of 10 sign languages across five language groups, this distribution becomes heavily skewed toward two-handed signs when looking specifically at a list of lexically plural concepts (i.e., concepts that are inherently plural). The motivation for this is argued to be that sign languages make use of *articulatory plurality*, which means that they map plural referents onto the plural articulators (e.g., the two hands) in an iconic manner. An informal glance at the distribution across languages reveals that while the list of random concepts shows a quite even distribution of one- vs. two-handed signs (Figure 6), the sampled lexically plural concepts show a two-handed preference across languages (Figure 5).

Setting the categorization of plural vs. random list aside, we also want to compare the proportion of two-handed signs to the perceived plurality of the individual concepts (section 2.3.2). Figure 7 shows the correlation between the plural ratings for individual concepts and the proportion of two-handed signs encoding the same concepts across languages. Here we see that there is a clear difference in patterning between plural and random items. Random items have generally low plural ratings (as expected), and are also evenly distributed across the y axis, demonstrating that non-plural items exhibit the expected 50/50 split between one- and two-handed signs. However,

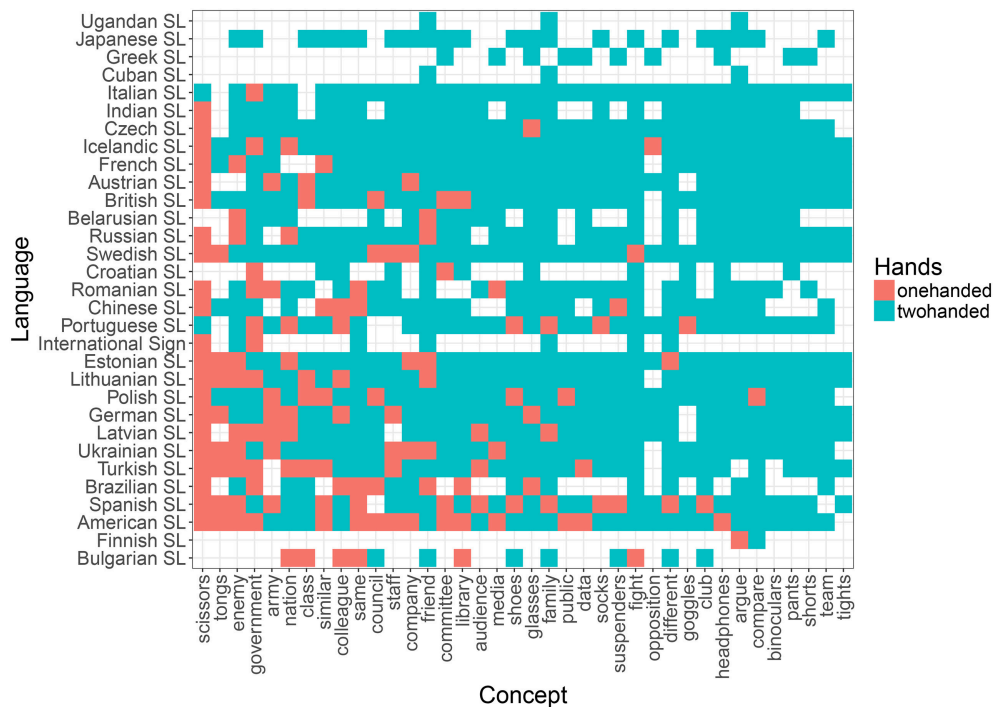


FIGURE 5 | Distribution of one- (red) vs. two-handed (blue) signs across concepts (x axis) and languages (y axis) in the lexical plural list.

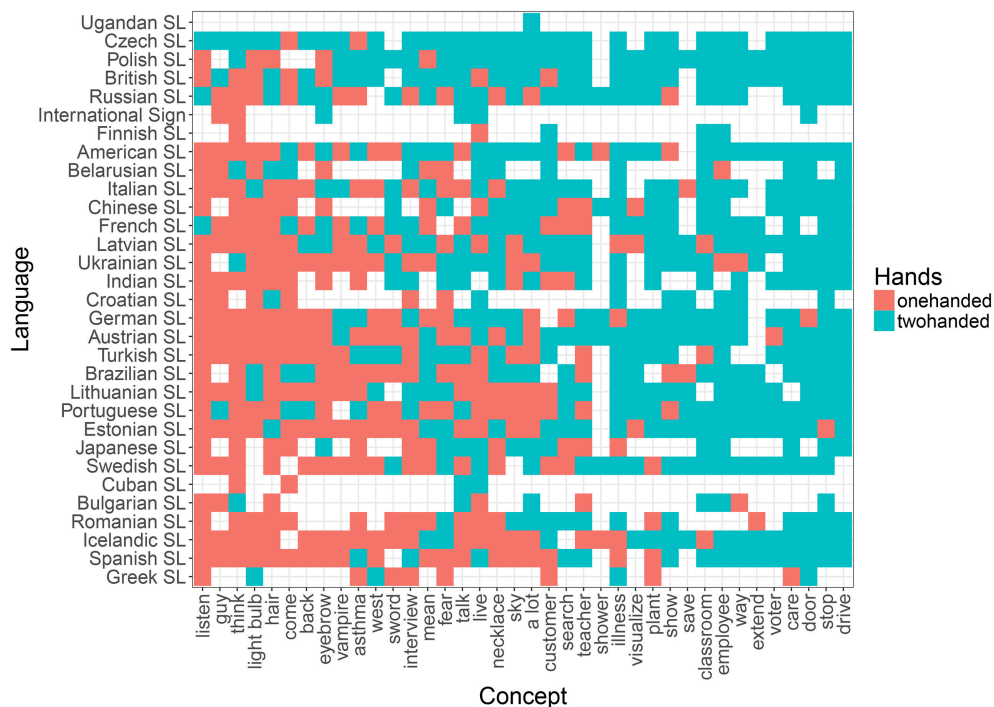


FIGURE 6 | Distribution of one- (red) vs. two-handed (blue) signs across concepts (x axis) and languages (y axis) in the random list.

the plural items have mixed plural ratings, although overall much higher than the random items, and are all clearly biased toward two-handed sign forms across languages. This gives

us a general visualization of plurality and two-handed forms being correlated. We have investigated this more rigorously using the model described in section 2.3.3. Its parameters were

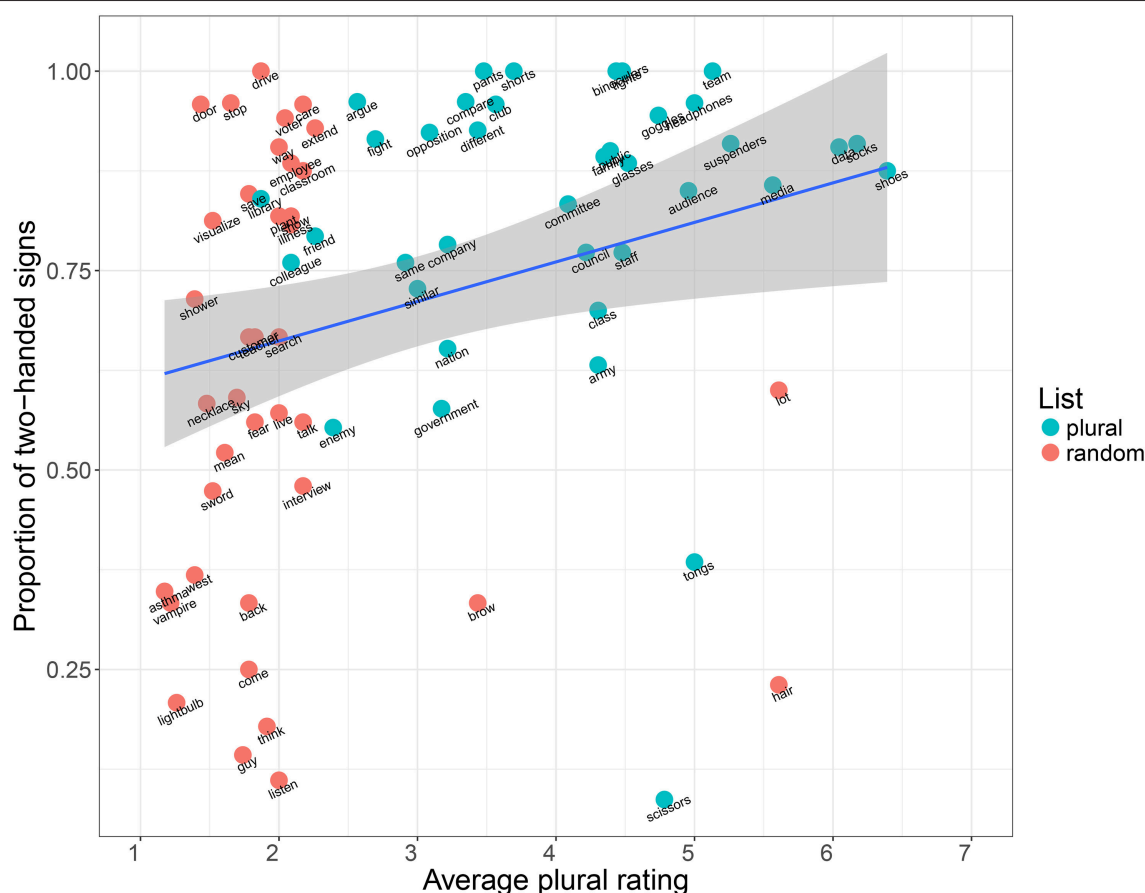


FIGURE 7 | Correlation between plurality ratings (x axis) and the proportion of two-handed signs across languages (y axis) for both plural (blue) and random (red) items.

estimated using the 81 concepts in the list ranked for lexical plurality described in section 2.3.2. **Table 5** summarizes some important parameters and their estimates. Due to the relatively low number of concepts studied, the posterior distributions of these parameters are fairly wide, but still allow us to draw a number of conclusions conditioned on the assumptions of our model:

- $P(\beta_p > 0) \approx 99.5\%$ and $P(\beta_p > 1) \approx 93.0\%$: lexical plurality is extremely likely to be a predictor of two-handed signs, and the effect is likely to be large.
- $P(\beta_f < 0) \approx 41.7\%$ and $P(|\beta_f| > 1) \approx 26.5\%$: we do not have evidence for frequency being a predictor of one-handed signs, but even a large effect (in either direction) can not be excluded.
- $P(\sigma^L < 0.67) \approx 97.5\%$: languages are fairly consistent in the overall distribution of one- and two-handed signs, and most variation is explained by other factors.
- $P(\sigma^C > 1.30) \approx 97.5\%$: concept-specific properties beyond lexical plurality are important predictors of one- or two-handedness. The α_c^C values at the bottom of **Table 5** provide some examples.

Looking at the individual concepts in the lexical plural list, we observe that all but two (“tongs” and “scissors”) are preferentially

encoded as two-handed signs across languages (**Figure 5**). This can not be explained by any of the other factors in our statistical model (see above, and **Table 5**). This is also reflected in the results of Börstell et al. (2016), for which “tongs” and “scissors” were the only concepts of those overlapping with this current study that were not encoded as two-handed signs in any sign language. A manual check in the *Spread the Sign* videos shows that “tongs” are mostly referred to by a handling depiction (i.e., showing how a hand uses tongs), whereas *all* languages have a one-handed depiction of “scissors” with the fingers representing the shears. As argued by Börstell et al. (2016), using dual/plural fingers for mapping plural referents is also an instance of articulatory plurality, only using a different individuation of articulators (i.e., fingers instead of hands) in the iconic mapping between articulators and plural referents. In **Figure 5**, we notice that there has been a misclassification of “scissors” as two-handed for two languages (Italian Sign Language and Portuguese Sign Language).

There are also some clear examples of the opposite tendency, where the other factors in our model are unable to explain two-handedness. For instance, the concepts DRIVE and DOOR are nearly universally two-handed signs in our sample, in spite of receiving low plurality rankings. They both have iconic

TABLE 5 | Estimates of some interesting model parameters, with credibility intervals centered at the median.

Parameter	Description	Percentile				
		2.5	25	50	75	97.5
β_p	Plurality regression coefficient	0.58	1.68	2.25	2.84	3.93
β_f	Frequency regression coefficient	-1.58	-0.40	0.18	0.77	1.92
a	Intercept	-1.02	-0.15	0.29	0.73	1.61
σ^C	Per-concept regression term <i>SD</i>	1.30	1.50	1.61	1.74	2.03
σ^L	Per-language regression term <i>SD</i>	0.20	0.33	0.41	0.49	0.67
σ^R	Per-rater bias term <i>SD</i>	0.62	0.75	0.83	0.93	1.17
σ^Q	Rating noise <i>SD</i>	1.47	1.50	1.52	1.53	1.57
$\alpha_{\text{SCISSORS}}^C$	Regression term of concept SCISSORS	-5.50	-4.36	-3.82	-3.32	-2.43
α_{HAIR}^C	Regression term of concept HAIR	-5.02	-4.09	-3.63	-3.19	-2.41
α_{GUY}^C	Regression term of concept GUY	-5.50	-4.04	-3.39	-2.81	-1.85
...						
α_{DOOR}^C	Regression term of concept DOOR	0.50	1.41	1.93	2.49	3.75
α_{SHORTS}^C	Regression term of concept SHORTS	0.22	1.33	1.98	2.71	4.38
α_{DRIVE}^C	Regression term of concept DRIVE	0.81	1.84	2.45	3.13	4.73

motivations, but concept-specific ones not captured by the notion of lexical plurality: driving a vehicle by holding the steering wheel (usually symmetrically, using both hands), and opening a door in a wall (asymmetric, with the non-dominant hand representing a reference point in the wall/door post). These are, however, motivations found by Lepic et al. (2016), for instance in having spatial configurations with one hand as the reference point to the other, such as in DOOR, or using plural hands to represent plural hands, as in DRIVE, although the meaning of “drive” is not inherently plural in itself (the plurality is in the limbs used for the activity).

3.2. Study 2: Sign Locations

For our second research question, we wanted to investigate whether there are cross-linguistically valid patterns of sign locations being iconic. Here we define iconicity based on the location ratings made by hearing non-signers asked to map concepts onto a body silhouette (see section 2.4.1). Using these ratings, we can also quantify the extent to which the cross-linguistic hand activity patterns align with non-signers' associations between concepts and the human body. **Table 6** shows that out of the six concepts from Börstell and Östling (2017), four are in the 4th percentile or lower, indicating a high degree of similarity to the location ratings. The remaining two concepts, FOOD and SAY, are both articulated at the mouth in nearly all of the languages in our sample. Thus, for these individual concepts, we find some clear examples of iconic locations across languages.

In order to visualize cross-linguistic similarity, **Figure 8** shows the hand activity across the languages in the *Spread the Sign* dictionary for the individual concepts. These concepts, also compared to the iconicity of locations task, have been chosen in part because they provide prototypical examples of the categories

TABLE 6 | Concepts and their similarity to location ratings (0 would indicate perfect agreement).

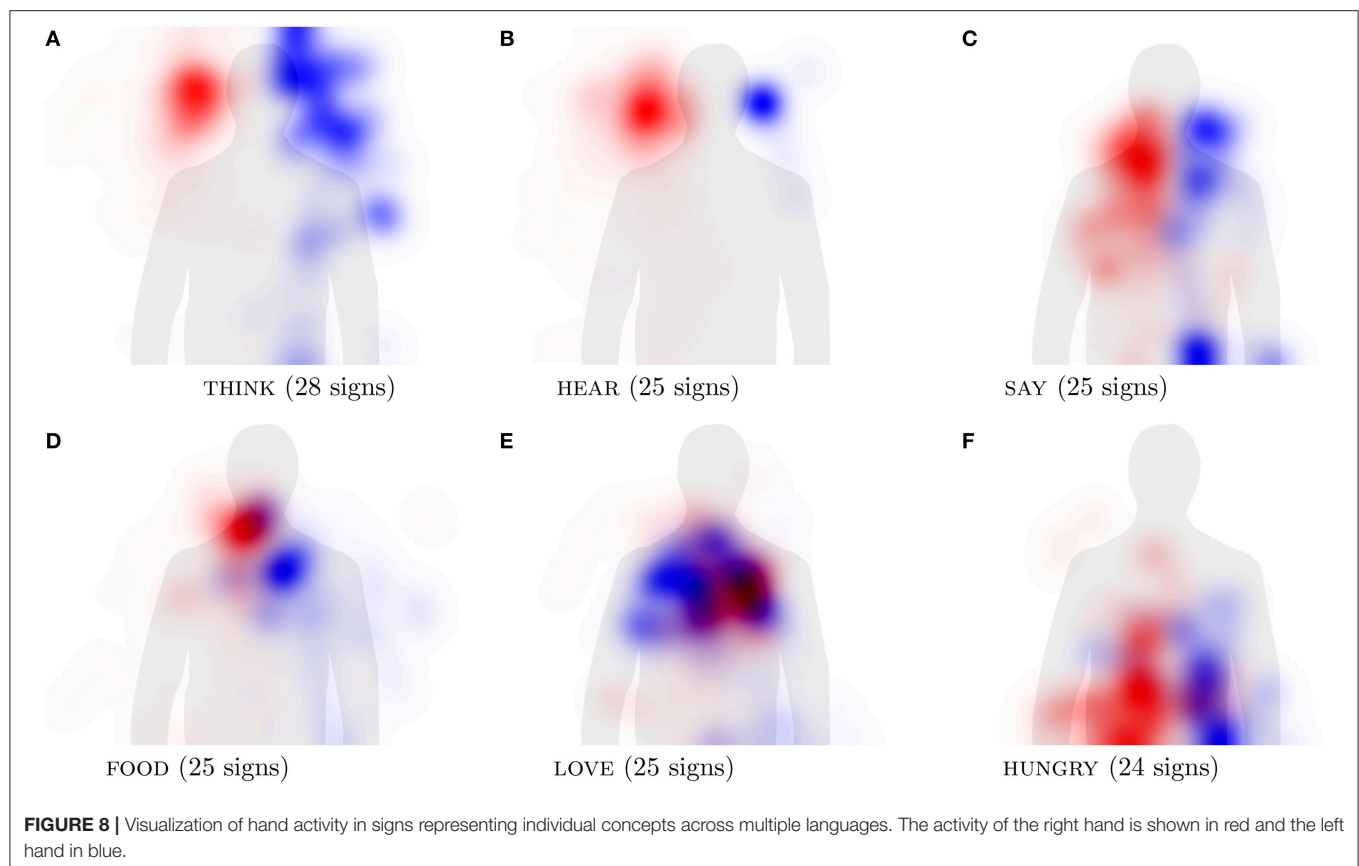
Concept	Concept only		All signs	
	Mean sim.	Percentile	Mean sim.	<i>SD</i>
FOOD	-0.84	26	-0.95	0.18
HEAR	-0.42	3	-1.03	0.32
HUNGRY	-0.37	1	-0.93	0.30
LOVE	-0.28	3	-0.52	0.18
SAY	-0.88	32	-1.02	0.28
THINK	-0.49	4	-1.10	0.33

Mean similarity scores are presented for signs in different languages representing the given concept (left) and for all signs in the Swadesh list (right).

investigated by Börstell and Östling (2017) for Swedish Sign Language. As expected, they show strong tendencies to be located in certain areas: THINK (forehead), HEAR (ears), SAY and FOOD (lower face), LOVE (chest, with crossed arms), HUNGRY (belly). These sign locations are clearly iconic in some sense, either directly or metaphorically: whereas “think,” “hear,” and “say” are all located at the body part directly involved in the respective activity (head/brain for thinking, ears for hearing, and mouth/throat for saying), “food” is located at the body part associated with a related action (mouth for eating). The concept “hungry” is located at the belly, in which hunger is felt, and “love” is located at the chest, which is explained by the metaphorical association of heart as the center of experiencing the emotion. In the visualization, some cases are clearer than others in that there is less cross-linguistic variation. For instance, the concept “hear” is clearly associated with the ears across languages, as is “think” with the (fore)head and “food” with the mouth. The concept “say” shows a slightly more variable location across languages, yet mostly centered around the mouth (see **Figure 8C**), as has previously been argued by Frishberg and Gough (2000, p. 117–118). This shows that iconic mappings between sign location and meaning—directly or metaphorically—are visible across languages, which is supported by the fact that they correlate with the locations identified as iconic for each concept by hearing non-signers (**Table 6**).

Rather than looking at individual concepts, **Figure 9** shows hand activity over larger categories with hundreds of individual signs each. The *Nouns* category is included as an example of overall hand activity in a semantically *non-coherent* group of concepts, while the remaining pictures in **Figure 9** represent semantically (relatively) coherent groups of concepts. Compare the result to Figures 3, 4 in Börstell and Östling (2017, p. 223), who used hand-coded data from Swedish Sign Language for similar categories.

In some cases the categories are strongly associated with particular body parts, which is reflected by our visualizations. For example, *Eyesight* (eyes), *Sound* (ears), and *Hair* (hair) are all clearly associated with the expected body location. In other cases, the categories are associated with large parts of the body, leading to less focused activity patterns, as in the *Clothes* and *Anatomy* categories. Overall, the activity areas are less distinct than for the



individual concepts shown in **Figure 8**, which is expected from the conflation of data from many signs across all languages into one image, but iconic mappings can still be observed.

Besides the figures shown here, we have generated visualizations of sign locations across languages for all individual concepts and sign categories present in the *Spread the Sign* database at the time of download¹⁰. This may prove useful for other researchers interested in further exploring the cross-linguistic patterns of iconicity with regard to sign locations and semantics.

4. DISCUSSION

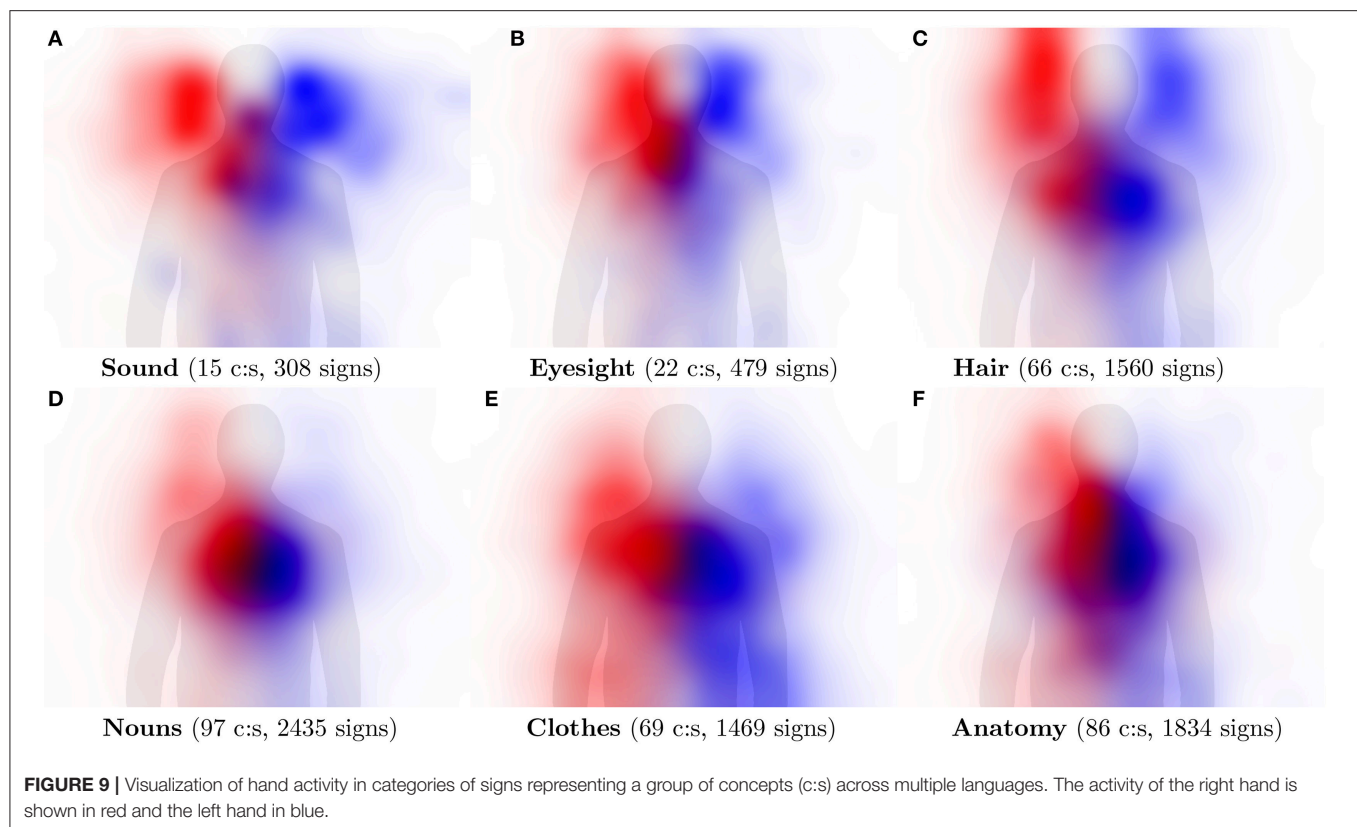
In this paper, we set out to explore the visual iconicity of signed languages, specifically targeting two phonological parameters—the number of articulators (specifically the number of hands) and the location used in a sign—investigated in the two separate studies presented here. The aim was to explore how these parameters relate to iconicity across languages. Furthermore, since the method of using automated visual processing of large-scale parallel sign language data is a novel one, we also wished to evaluate the accuracy and usefulness of this method, for this and future studies.

¹⁰<https://www.ling.su.se/sign-language-iconicity>

4.1. Number of Articulators

For our first study, we wanted to investigate the distribution of one- vs. two-handed signs across the sign languages in the dataset. Previous studies have suggested that sign language lexicons exhibit quite an even split between one- and two-handed signs (Börstell et al., 2016; Crasborn and Sáfár, 2016; Lepic et al., 2016). Our statistical model (section 3.1.3), using 81 concepts for which we have lexical plurality ratings, indicates that none of the sign languages in our sample has a strong preference for either one- or two-handed signs. Sampling signs corresponding to the concepts in a Swadesh list, which may be used as an approximation of core vocabulary, we see that the distribution corresponds to the one predicted, with an even distribution across all signs and languages. Looking at the individual languages in our data, there is one exception to the overall 50/50 distribution that clearly stands out: Czech Sign Language with 90.1% two-handed signs. This result, along with those of the second and third most two-handed languages (British Sign Language and Russian Sign Language), are likely due to the systematic errors in the automatic processing of those languages described in section 3.1.1. For this reason they have been excluded from the statistical analysis, along with Portuguese Sign Language which had similar problems.

However, when expanding our analysis to a much larger set of concepts, using all concepts encoded by a single word in English to avoid phrases and complicated multisign constructions, we



see that the proportion of two-handed signs notably increases to 71% of all signs across languages. Although we do not have an answer to why this is the case, it is possible that less frequent signs are more likely to be two-handed, if one assumes that the transition from two-handed to one-handed could be a frequency-induced economically motivated reduction (cf. Crasborn and Sáfár, 2016, p. 244). The frequency hypothesis was tested in our statistical model using 81 concepts, but could neither be supported nor conclusively rejected. It may also simply be the case that the sample, though consisting of simplex word forms in English, are encoded by complex, depicting, or multi-sign constructions to a larger extent than the smaller core vocabulary set, as larger lexical databases for individual sign languages—hence implying both frequent and infrequent lexical items included—have previously been found to adhere to the 50/50 pattern, too (see Börstell et al., 2016, p. 393). To what extent the two-handed prominence is a result of misclassification by the algorithm is not known, but manual checks of the classifications have shown mistakes in both directions, without a clear bias in either direction. Nonetheless, the previous claims of a 50/50 distribution are confirmed for a core vocabulary sample, but less so for all languages and concepts in a larger sample of signs.

We do not have a definitive answer as to why the overall distribution of signs exhibits a 50/50 split. It is possible that it has to do with a general preference to organize linguistic units as maximally distinct forms, in a similar fashion to spoken language phoneme inventories preferring distinct (physically

and acoustically dispersed) vowels (Lindblom, 1986; Vaux and Samuels, 2015). With this analogy, sign languages would use the one- vs. two-handed division as one main distinction in phonological form, and without taking articulatory economy into account, this could generate an even split. Since we see that at least two-handed forms are associated with certain semantics, we still expect systematicity (and, in our case with plurality, *iconicity*) in some parts of the lexicon, which affects this distribution locally, and further investigations into form–meaning patterns may resolve this issue. As argued for spoken languages (see Dingemanse et al., 2015), there is an advantage in balancing non-arbitrariness and arbitrariness in the word forms of a language, since they have different benefits. For example, whereas the vowel /i/ may be used to iconically denote “smallness,” it would be a major restriction for a language to require such a form–meaning mapping for all uses of /i/—that is, the /i/ would be restricted to a very limited set of words with certain semantics. Similarly, if the one- vs. two-handed division is a useful phonological distinction, it would be a severe limitation to not let that distinction be used arbitrarily. This does, however, not entail that it can *never* be used iconically, and we argue that two-handed forms are often iconically mapped onto plural meaning when possible, but plurality is not a strict requirement when forming two-handed signs.

Concerning the association between lexical plurality and two-handed signs, as found by Börstell et al. (2016), we can say with certainty that this is a valid claim across the

languages in our dataset. Our statistical model showed that concepts are significantly more likely to be encoded by a two-handed form if the concept carries plural semantics, even when taking language- and concept-specific variation as well as frequency into account. We argue that this is not only a systematic, but an iconic mapping between plurality and two-handed forms, and an instance of *articulatory plurality* (Börstell et al., 2016). We further corroborate that this is a cross-linguistically valid pattern, and a unique iconic feature of signed language employing the visual modality—unique in the sense of having multiple symmetrical articulators available simultaneously, which adds another articulatory dimension to the previously known mapping between quantity of form and quantity of meaning (Lakoff and Johnson, 1980; Dingemanse et al., 2015). Thus, whereas the MORE OF FORM IS MORE OF CONTENT metaphor here limits form to the two hands, the mapping is found in reduplication among spoken languages, and sign languages may also use other multiple articulators for this mapping (e.g., individuated fingers). However, we do not assume the use of multiple articulators in this way to be a unique property of signed language, but rather unique to the visual modality, as it has been shown that hearing non-signers too adhere to articulatory plurality with regard to number of hands when inventing silent gestures for lexically plural concepts (Börstell et al., 2016).

4.2. Sign Locations

For our second study, we wanted to see to what extent sign locations are employed in similar iconic mappings across languages. In a previous study, Börstell and Östling (2017) showed that it is possible to visualize the location–meaning association of lexical signs in a quantified manner for a single sign language (Swedish Sign Language). By using a series of automated methods for identifying hand activity, including extrapolating hand positions and normalizing sign locations with body pose joint locations as reference points, we were able to visualize areas of hand activity across languages for both individual concepts and larger lexical or semantic categories. We have also demonstrated that these location–meaning associations are iconic, since they correlate strongly with the assignment of iconic locations to concept meanings resulting from an experimental task involving hearing non-signers. Some of these mappings are concretely iconic, such as associating “think” with the (fore)head and “say” with the mouth, and others are metaphorically iconic, such as using the chest/heart area to represent “love.” We thus conclude that both direct and metaphorical mappings contribute to iconic sign formation across sign languages, as argued by Taub (2001). However, we cannot based on our data quantify the extent to which sign language lexicons are iconically motivated, only that the iconic motivation is an available strategy and one that can be verified with quantitative methods.

For the concept categories, the picture becomes more indistinct. This is unsurprising considering it is an overlay of hundreds—or even thousands—of signs corresponding to the concepts in the category, visualized in a single image. Nonetheless, we do see that a category such as *Sound* shows

a location pattern similar to that of the individual concept HEAR—i.e., around the ears—albeit with less distinct boundaries and more noise. Thus, we still consider our method useful in investigating cross-linguistic patterns of systematicity and iconicity in sign locations, which are indeed visible in several cases. It is possible that the visualization across languages and concepts simultaneously would show clearer patterns if one were to sample individual concepts based on pre-defined semantic features—that is, rather than by the crude categorizations available in *Spread the Sign*—but it is still likely that some discrepancies and noise will be present. Even though this and a plethora of previous studies have shown that iconicity is a clearly visible feature of signed languages, we would not expect an exact matching of sign locations across languages, even when the location is iconic, seeing as different signs may draw on different iconic mappings (cf. Klima and Bellugi, 1979; Lepic et al., 2016), and thus some variability will appear as noise in this type of visualization. Furthermore, as we argued concerning the one- vs. two-handed division, we would not expect all sign locations to be motivated by meaning, since that could induce unwanted restrictions on possible forms. However, we do acknowledge that many signs are in fact partially iconic, although the iconicity may be found in other parameters than specifically location. Some claims have been made for sign locations exhibiting systematicity rather than iconicity, for instance with signs articulated on the nose or the chin being associated with negative or derogatory meanings in American Sign Language (Frishberg and Gough, 2000, p. 116–118). Thus, we assume that the benefits of balancing arbitrariness and non-arbitrariness are as relevant to signed languages as they are to spoken languages, even if the proportions between the two strategies may differ.

4.3. Evaluating the Method

The main reason for using automated analysis is efficiency: we can perform several types of analysis on hundreds of thousands of individual signs, in dozens of languages, much faster than any manual coding. This allows large-scale analyses that would be unrealistic to perform manually, and furthermore creates a great potential for exploratory work since working hypotheses can be evaluated in hours rather than months. However, since the computer models available currently perform at less than human performance, there are certain limitations to what we are able to do.

One obvious limitation in using automated classification of sign locations using visual video processing is the fact that this classification only concerns two-dimensional space. The visual modality uses three dimensions, hence any sign location is three-dimensional. Thus, while the automated classification may provide us with height and width information of signs, it does not provide us with any information about the depth of the articulation, such as whether the sign is articulated close to or far away from the signer’s body, nor whether the sign has contact with the body part in front of which it is articulated. It is known from previous work on sign iconicity that contact as part of the articulation at a location is in itself meaningful (e.g., Taub, 2001; Meir et al., 2013), but this information is currently unavailable to us. This means, for instance, that signs articulated in neutral

space *in front of* the signer are conflated with signs articulated *on* the signer's chest/torso, and whereas the former may lack an iconic mapping between meaning and location, the latter may very well feature such a mapping (e.g., EMOTION being centered around the signer's chest). It is possible that a model that could single out only signs with body contact articulation would show even stronger patterns of iconicity in different locations.

The above-mentioned limitations are inherent to the video analysis method we use, but we would also like to stress that unexpected systematic and non-systematic errors can be numerous in some cases, even with an automatic system that has a high overall accuracy. For instance, seemingly trivial differences in lightning conditions and how the frame was centered forced us to discard data from four languages. This was discovered only after manually annotating a random subset of signs from each language as one-handed or two-handed. Such an annotation also gives bounds on the rate of non-systematic errors, which should ideally be low compared to the agreement between human raters. We thus strongly encourage researchers to annotate as much data as is practical manually for validation purposes, both to quantify expected problems (when relevant) and to discover unexpected ones. In our case we only considered one variable—the number of active hands—but future work using more fine-grained distinctions, such as handshape and non-manual signals, should ideally contain an evaluation of the accuracy with which each of the variables can be automatically extracted. This is sometimes difficult to do in case the estimated quantity would be complex to annotate, such as in our Study 2 in which we estimate overall levels of hand activity. In these cases, proxy measures such as sign locations may have to be used, although such values are categorical based on phonological assumptions.

Nonetheless, our method here has proven useful in detecting systematic form–meaning mappings across language and concepts even without such an adjustment, showing that automated video processing methods can be useful in analyzing large datasets across languages in the visual modality.

4.4. Conclusions

In this paper, we have demonstrated that computational methods may be used in order to detect and quantify patterns of systematicity between form and meaning in the visual language modality. By comparing these patterns to iconicity measurements, we have also been able to show that several of these systematic associations are in fact iconic in nature. For the articulators, we corroborate previous findings, here with a much larger language sample, that lexical plurality of concepts

(whether categorical or scalar) correlates with the likelihood of sign languages using a two-handed form: plurality of meaning is iconically mapped onto plural articulators. For sign locations, we see that there are systematic patterns visible across languages, and that these correlate with iconicity ratings given by hearing non-signers, suggesting that concrete and metaphorical iconicity is employed across our sampled sign languages.

Not only do our findings provide further proof of the iconicity potential of language, here specifically the visual iconicity prevalent across languages in the signed modality, but it also adds to a growing body of quantitative linguistics, including research investigating non-arbitrariness across large sets of words and languages (e.g., Urban, 2011; Blasi et al., 2016). We specifically show that the number of hands and the location of signs may be used iconically, across sign languages. We believe that this is one further step toward measuring the prevalence of (non-)arbitrariness in human language, not only across languages but also across modalities, which is an important task in order to explore the fundamental properties that constitute linguistic structure. We also believe that computational methods may become increasingly useful in quantifying language in the visual modality, such as extracting formational features of signs or gestures (e.g., handshapes, movement volume, and manner). To some extent, this has been done previously, but then normally already at the data collection stage, by using technology such as Motion Capture (e.g., Puupponen et al., 2015) or Kinect (e.g., Namboodiripad et al., 2016). Here, we show that even pre-collected video data may be analyzed computationally, in post-production, by employing automated video processing methods.

AUTHOR CONTRIBUTIONS

The project idea and study design were devised by CB and RÖ (equal contribution). Plurality and location iconicity ratings were collected by CB and RÖ. Sign language data collection and processing were carried out by SC and RÖ (equal contribution), with assistance from CB. The manuscript was drafted by CB and RÖ (equal contribution), and reviewed by SC.

ACKNOWLEDGMENTS

We wish to thank the reviewers and editors for comments and suggestions on the paper, as well as Thomas Hörberg for some statistical advice and Hedvig Skirgård for literature suggestions. We also wish to thank the respondents of our plurality rating questionnaire and location rating task. The Titan X Pascal used for this research was donated by the NVIDIA Corporation.

REFERENCES

- (2012). *Spreadthesign*. European Sign Language Center.
- Acquaviva, P. (2008). *Lexical Plurals*. New York, NY: Oxford University Press.
- Acquaviva, P. (2016). Structures for plurals. *Lingvist. Invest.* 39, 217–233. doi: 10.1075/li.39.2.01acq
- Attarde, I. P. (2007). *Encyclopedic Graded Grammar, Vol. 1*. Lulu. Available online at: <http://www.lulu.com/>
- Auracher, J. (2017). Sound iconicity of abstract concepts: place of articulation is implicitly associated with abstract concepts of size and social dominance. *PLoS ONE* 12:e0187196. doi: 10.1371/journal.pone.0187196
- Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., and Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proc. Natl. Acad. Sci. U.S.A.* 113, 10818–10823. doi: 10.1073/pnas.1605782113

- Börstell, C., Hörberg, T., and Östling, R. (2016). Distribution and duration of signs and parts of speech in Swedish Sign Language. *Sign. Lang. Linguist.* 19, 143–196. doi: 10.1075/sll.19.2.01bor
- Börstell, C., Lepic, R., and Belsitzman, G. (2016). “A show of hands: plurality as a feature of two-handed forms in silent gesture,” in *7th Conference of the International Society for Gesture Studies (ISGS 7)* (Paris: Sorbonne Nouvelle University).
- Börstell, C., Lepic, R., and Belsitzman, G. (2016). Articulatory plurality is a property of lexical plurals in sign language. *Lingvist. Invest.* 39, 391–407. doi: 10.1075/li.39.2.10bor
- Börstell, C., and Östling, R. (2017). “Iconic locations in Swedish Sign Language: mapping form to meaning with lexical databases,” in *Proceedings of the 21st Nordic Conference on Computational Linguistics (NODALIDA 2017)*, NEALT Proceedings Series 29, ed J. Tiedemann (Gothenburg: Linköping Electronic Press), 221–225.
- Brennan, M. (1990). *Word Formation in British Sign Language*. Doctoral dissertation, Stockholm University.
- Brennan, M. (2005). Conjoining word and image in British Sign Language (BSL): an exploration of metaphorical signs in BSL. *Sign. Lang. Stud.* 5, 360–382. doi: 10.1353/sls.2005.0007
- Brentari, D. (ed.). (2010). *Sign Languages: A Cambridge Language Survey*. New York, NY: Cambridge University Press.
- Cao, Z., Simon, T., Wei, S., and Sheikh, Y. (2016). Realtime multi-person 2d pose estimation using part affinity fields. *CoRR* abs/1611.08050. *arXiv:1611.08050*.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: a probabilistic programming language. *J. Stat. Softw.* 76, 1–32. doi: 10.18637/jss.v076.i01
- Cormier, K., Fenlon, J., and Schembri, A. (2015). Indicating verbs in British Sign Language favour motivated use of space. *Open Linguist.* 1, 684–707. doi: 10.1515/opli-2015-0025
- Crasborn, O., and Sáfár, A. (2016). “An annotation scheme to investigate the form and function of hand dominance in the Corpus NGT,” in *A Matter of Complexity*, eds R. Pfau, M. Steinbach, and A. Herrmann (Boston, MA; Berlin; Preston: De Gruyter Mouton & Ishara Press), 231–251.
- Cuskley, C., Simner, J., and Kirby, S. (2017). Phonological and orthographic influences in the bouba–kiki effect. *Psychol. Res.* 81, 119–130. doi: 10.1007/s00426-015-0709-2
- de Saussure, F. (1916). *Cours de Linguistique Générale*. Paris: Payot.
- Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Linguist. Lang. Compass* 6, 654–672. doi: 10.1002/lnc3.361
- Dingemanse, M., Blasi, D. E., Lupyán, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends Cogn. Sci.* 19, 603–615. doi: 10.1016/j.tics.2015.07.013
- Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., and Mitterer, H. (2016). What sound symbolism can and cannot do: testing the iconicity of ideophones from five languages. *Language* 92, e117–e133. doi: 10.1353/lan.2016.0034
- Dudis, P. G. (2004). Body partitioning and real-space blends. *Cogn. Linguist.* 15, 223–238. doi: 10.1515/cogl.2004.009
- Fort, M., Lammertink, I., Peperkamp, S., Guevara-Rukoz, A., Fikkert, P., and Tsuji, S. (2018). Symbouki: a meta-analysis on the emergence of sound symbolism in early language acquisition. *Dev. Sci.* e12659. doi: 10.1111/desc.12659. [Epub ahead of print].
- Frishberg, N., and Gough, B. (2000). Morphology in American Sign Language. *Sign. Lang. Linguist.* 3, 103–131. doi: 10.1075/sll.3.1.08fri
- Grose, D., Wilbur, R. B., and Schalber, K. (2007). Events and telicity in classifier predicates: a reanalysis of body part classifier predicates in ASL. *Lingua* 117, 1258–1284. doi: 10.1016/j.lingua.2005.06.014
- Hammarström, H., Forkel, R., and Haspelmath, M. (2017). *Glottolog 3.0*. Jena: Max Planck Institute for the Science of Human History. Available online at: <http://glottolog.org/>
- Haspelmath, M. (2007). “Further remarks on reciprocal constructions,” in *Reciprocal Constructions*, ed V. Nedjalkov (Amsterdam: Benjamins), 2087–2115.
- Hockett, C. F. (1960). The origin of speech. *Sci. Am.* 203, 89–96.
- Jepsen, J. B., De Clerck, G., Lutalo-Kiingi, S., and McGregor, W. B., (eds.). (2015). *Sign Languages of the World: A Comparative Handbook*. Berlin; Boston, MA; Preston: De Gruyter & Ishara Press.
- Kimmelman, V., Klezovich, A., and Moroz, G. (2018). *Iconicity Patterns in Sign Languages*. Available online at: https://sl-iconicity.shinyapps.io/iconicity_patterns/
- Klima, E. S., and Bellugi, U. (1979). “Iconicity in signs and signing,” in *The Signs of Language*, eds E. S. Klima and U. Bellugi (Cambridge, MA: Harvard University Press), 9–34.
- Köhler, W. (1929). *Gestalt Psychology*. New York, NY: Liveright.
- Koptjevskaja-Tamm, M. (2004). “Mass and collection,” in *Morphologie: ein Internationales Handbuch zur Flexion und Wortbildung = Morphology Vol. 2*, eds G. Booij, J. Mugdan, and C. Lehmann (Berlin: Walter de Gruyter), 1067–1072.
- Kwon, N., and Round, E. R. (2015). Phonaesthemes in morphological theory. *Morphology* 25, 1–27. doi: 10.1007/s11525-014-9250-z
- Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.
- Lauwers, P., and Lammert, M. (2016). Introduction. *Lingvist. Invest.* 39, 207–216. doi: 10.1075/li.39.2.001int
- Leech, G., Rayson, P., and Wilson, A. (2001). *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Longman. Available online at: <http://ucler.lancs.ac.uk/bncfreq/>
- Lepic, R. (2015). *Motivation in Morphology: Lexical patterns in ASL and English*. Doctoral dissertation, University of California, San Diego, CA.
- Lepic, R., Börstell, C., Belsitzman, G., and Sandler, W. (2016). Taking meaning in hand: iconic motivations for two-handed signs. *Sign. Lang. Linguist.* 19, 37–81. doi: 10.1075/sll.19.1.02lep
- Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., et al. (2014). Microsoft COCO: common objects in context. *CoRR* abs/1405.0312. doi: 10.1007/978-3-319-10602-1_48
- Lindblom, B. (1986). “Phonetic universals in vowel systems,” in *Experimental Phonology*, eds J. J. Ohala and J. J. Jaeger (Orlando, FL: Academic Press), 13–44.
- Lockwood, G., and Dingemanse, M. (2015). Iconicity in the lab: a review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Front. Psychol.* 6:1246. doi: 10.3389/fpsyg.2015.01246
- Lockwood, G., Hagoort, P., and Dingemanse, M. (2016). How iconicity helps people learn new words: neural correlates and individual differences in sound-symbolic bootstrapping. *Collabra* 2:7. doi: 10.1525/collabra.42
- Meir, I. (2010). Iconicity and metaphor: constraints on metaphorical extension of iconic forms. *Language* 86, 865–896. doi: 10.1353/lan.2010.0044
- Meir, I., Padden, C., Aronoff, M., and Sandler, W. (2013). Competing iconicities in the structure of languages. *Cogn. Linguist.* 24, 309–343. doi: 10.1515/cog-2013-0010
- Meir, I., and Tkachman, O. (2018). “Iconicity,” in *Oxford Research Encyclopedia of Linguistics*, ed M. Aronoff (Oxford: Oxford University Press).
- Mihatsch, W. (2016). Collectives, object mass nouns and individual count nouns. *Lingvist. Invest.* 39, 289–308. doi: 10.1075/li.39.2.05mih
- Monaghan, P., Shillcock, R. C., Christiansen, M. H., and Kirby, S. (2014). How arbitrary is language? *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130299. doi: 10.1098/rstb.2013.0299
- NambooDiripad, S., Lenzen, D., Lepic, R., and Verhoef, T. (2016). Measuring conventionalization in the manual modality. *J. Lang. Evol.* 1, 109–118. doi: 10.1093/jole/lzw005
- Occhino, C., Anible, B., Wilkinson, E., and Morford, J. P. (2017). Iconicity is in the eye of the beholder: how language experience affects perceived iconicity. *Gesture* 16, 100–126. doi: 10.1075/gest.16.1.04occ
- Occhino, C., and Wilcox, S. (2016). Constructing signs: place as a symbolic structure in signed languages. *Cogn. Linguist.* 27, 371–404. doi: 10.1515/cog-2016-0003
- Ortega, G. (2017). Iconicity and sign lexical acquisition: a review. *Front. Psychol.* 8:1280. doi: 10.3389/fpsyg.2017.01280
- Padden, C., Meir, I., Hwang, S.-O., Lepic, R., Seegers, S., and Sampson, T. (2013). Patterned iconicity in sign language lexicons. *Gesture* 13, 287–308. doi: 10.1075/gest.13.3.03pad
- Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol.* 1:227. doi: 10.3389/fpsyg.2010.00227
- Pfau, R., and Steinbach, M. (2003). Optimal reciprocals in German Sign Language. *Sign. Lang. Linguist.* 6, 3–42. doi: 10.1075/sll.6.1.03pfa

- Pfau, R., and Steinbach, M. (2006). Pluralization in sign and in speech: a cross-modal typological study. *Linguist. Typol.* 10, 135–182. doi: 10.1515/LINGTY.2006.006
- Pfau, R., and Steinbach, M. (2016). Modality and meaning: plurality of relations in German Sign Language. *Lingua* 170, 69–91. doi: 10.1016/j.lingua.2015.11.002
- Poggi, I. (2008). Iconicity in different types of gestures. *Gesture* 8, 45–61. doi: 10.1075/gest.8.1.05pog
- Puupponen, A., Wainio, T., Burger, B., and Jantunen, T. (2015). Head movements in Finnish Sign Language on the basis of Motion Capture data: a study of the form and function of nods, nodding, head thrusts, and head pulls. *Sign Lang. Linguist.* 18, 41–89. doi: 10.1075/sll.18.1.02puu
- Schmidtke, D. S., Conrad, M., and Jacobs, A. M. (2014). Phonological iconicity. *Front. Psychol.* 5:80. doi: 10.3389/fpsyg.2014.00080
- Taub, S. F. (2001). *Language From the Body: Iconicity and Metaphor in ASL*. Cambridge: Cambridge University Press.
- Thompson, R. L., Vinson, D. P., Woll, B., and Vigliocco, G. (2012). The road to language learning is iconic: evidence from British sign language. *Psychol. Sci.* 23, 1443–1448. doi: 10.1177/0956797612459763
- Urban, M. (2011). Conventional sound symbolism in terms for organs of speech: a cross-linguistic study. *Folia Linguist.* 45, 199–214. doi: 10.1515/flin.2011.007
- van der Hulst, H. (1996). On the other hand. *Lingua* 98, 121–143. doi: 10.1016/0024-3841(95)00035-6
- Vaux, B., and Samuels, B. (2015). Explaining vowel systems: dispersion theory vs natural selection. *Linguist. Rev.* 32, 573–599. doi: 10.1515/tlr-2014-0028
- Winter, B., Perlman, M., Perry, L. K., and Lupyan, G. (2017). Which words are most iconic?: Iconicity in English sensory words. *Interact. Stud.* 18, 430–451. doi: 10.1075/is.18.3.07win
- Wisniewski, E. J. (2010). “On using count nouns, mass nouns, and pluralia tantum: what counts?” in *Kinds, Things, and Stuff: Mass Terms and Generics*, ed F. J. Pelletier (New York, NY: Oxford Scholarship Online), 1–24.
- Zwitserslood, I., Perniss, P., and Asli, Ö. (2012). An empirical investigation of expression of multiple entities in Turkish Sign Language (TID): considering the effects of modality. *Lingua* 122, 1636–1667. doi: 10.1016/j.lingua.2012.08.010

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Östling, Börstell and Courtaux. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Automating the Production of Communicative Gestures in Embodied Characters

Brian Ravenet^{1*}, Catherine Pelachaud², Chloé Clavel¹ and Stacy Marsella³

¹ Laboratoire Traitement et Communication de l'Information (LTCI), Télécom ParisTech, Paris, France, ² Centre National de la Recherche Scientifique, Institut des Systèmes Intelligents et Robotiques, Sorbonne University, Paris, France, ³ College of Computer and Information Science (CCIS), Northeastern University, Boston, MA, United States

In this paper we highlight the different challenges in modeling communicative gestures for Embodied Conversational Agents (ECAs). We describe models whose aim is to capture and understand the specific characteristics of communicative gestures in order to envision how an automatic communicative gesture production mechanism could be built. The work is inspired by research on how human gesture characteristics (e.g., shape of the hand, movement, orientation and timing with respect to the speech) convey meaning. We present approaches to computing where to place a gesture, which shape the gesture takes and how gesture shapes evolve through time. We focus on a particular model based on theoretical frameworks on metaphors and embodied cognition that argue that people can represent, reason about and convey abstract concepts using physical representations and processes, which can be conveyed through physical gestures.

Keywords: metaphorical gestures, embodied conversational agents, communicative behaviors, text analysis, embodied cognition

OPEN ACCESS

Edited by:

Wendy Sandler,
University of Haifa, Israel

Reviewed by:

Alan Cienki,
VU University Amsterdam,
Netherlands
Michael Neff,
University of California, Davis,
United States

*Correspondence:

Brian Ravenet
brian.ravenet@isir.upmc.fr;
brian.ravenet@u-psud.fr

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 15 December 2017

Accepted: 14 June 2018

Published: 09 July 2018

Citation:

Ravenet B, Pelachaud C, Clavel C and
Marsella S (2018) Automating the
Production of Communicative
Gestures in Embodied Characters.
Front. Psychol. 9:1144.
doi: 10.3389/fpsyg.2018.01144

1. INTRODUCTION

Today, computers are essential in a wide range of activities, from solving mathematical problems to mediating our social interactions. Leveraging growth in computational power and functionality, researchers in the field of Embodied Conversational Agents (ECAs) aim to develop computer systems that can engage users in natural interactions. ECAs are virtual characters, usually with human-like appearances, endowed with the ability to use natural language and nonverbal behaviors the same way humans would (Cassell, 2000). They can be used as pedagogical assistants (Harvey et al., 2015), video-game characters (Gris et al., 2016) or they can also be integrated in more complex social simulations for medical purposes (Lisetti et al., 2015). Because their effectiveness relies on the user interacting with them the same way she would with another human, ECAs need to be able to decode and reproduce complex human communicative signals. While using only verbal communication may be satisfying for inputting basic commands, face-to-face communication requires the combination of speech with nonverbal behaviors that allows other communicative functions to be expressed simultaneously. For instance, research has highlighted how nonverbal behaviors are used by humans for disambiguation, clarification (Calbris, 2011), turn-taking management (Duncan, 1972) and socio-emotional expression while talking (Argyle, 1972). Therefore, in order to develop richer and more efficient natural interactions, ECAs require not only verbal capabilities but nonverbal ones as well.

While many communicative functions and nonverbal behaviors could be addressed, in this paper we focus on representational gestures which are gestures used to accompany and illustrate the

content of the speech. In particular, we present an approach to automatically producing metaphoric gestures that are aligned with the speech of the agent in terms of timing and meaning (Cienki and Müller, 2008).

Metaphoric gestures use the physical behavior of a gesture, its form and motion, to convey abstract concepts. For example, although ideas are immaterial, a gesture that is a sideways flip of the hand can convey the speaker's rejection of an idea, as if an idea is a physical object with physical features such as form and location and therefore it can be held and discarded. This view is in line with the embodied cognition theories that argue that the same set of sensory and motor representations we use to make sense of and act in our world are also used to make sense of, reason and communicate about abstract concepts (Barsalou, 1999; Kendon, 2000; Tversky and Hard, 2009). Thought, and the message to convey, is therefore construed in terms of concrete elements, the properties of those elements and actions on them. In this way, an "idea" conceptualized as a concrete object possesses physical properties, such as size, location or weight, that are tied to the abstract properties. For example, an important idea is an idea that is big in size, ideas can be thrown away, etc. Beyond offering a physical representation to abstract elements, embodied cognition considers that reasoning and thought processing are actions taken on these representations (Johnson-Laird, 2006), and that gestures, in particular metaphoric gestures, are physical representations of these actions realized at the conceptual level (Hostetter and Alibali, 2008, 2010). In other words, holding an idea in our hand or rejecting it by a sideways flip of the hand is a mirroring of actions taken at the conceptual level, in effect, considering an idea to examine or dismiss it.

This work explores theoretical frameworks on metaphors and on how people represent and transfer physical properties from one concept to another that have been highlighted by researchers in the field of embodied cognition (Wilson and Golonka, 2013). We aim at capturing and understanding the specific characteristics of communicative gestures in order to envision how an automatic communicative gesture production mechanism, inspired by these theoretical foundations on human embodied cognition and on related work, could be built. Gesture characteristics (e.g., shape of the hand, movement, orientation or timing with respect to the speech) should convey the desired meaning. A system capable of producing automatically relevant and meaningful gestures is of particular interest for ECAs as they often rely on canned templates or on scripted scenarios. Due to the growing popularity of procedurally generated content in virtual worlds, a system that can control autonomously the verbal and the nonverbal behaviors of virtual characters could be used in a variety of applications, from video games and movie tools to virtual assistants. Our work faces the following challenges: identifying a common representation between speech and gestures that could be computationally manipulated, proposing a mechanism to extract semantic elements of this representation from the speech of the agent, associating these elements to gesture characteristics and finally combining these gesture characteristics to align them with the speech of the agent. Throughout this article, we detail the different conceptual components of our

architecture and also their preliminary implementations to demonstrate the feasibility of such a system. While we aim for a balanced description of each of the conceptual components, some of them are more advanced in terms of implementation and will have a higher level of technical detail.

This article is organized as follows. In section 2, we present the theoretical foundations of our study on gestures, embodied cognition and discourse. We accompany this review with a discussion on the challenges that arise from replicating these human phenomena within an ECA. In section 3, we review and analyze existing solutions that tackled parts of our challenge. We leverage this literature to propose a system capable of generating metaphoric gestures starting from the text to be said by the virtual agent (see section 4). Finally, in section 5, we discuss the limits and perspectives of our approach and outline the requirements for future evaluations.

2. GESTURES AND MEANINGS

While talking, humans produce various nonverbal behaviors that accompany the discourse. Among these behaviors, communicative gestures can carry different meanings. They can illustrate an idea, mimic an action or the shape of an object, indicate a point in space or even mark an emphasis (McNeill, 1992). Various taxonomies of gestures have been proposed to encompass these varieties of meaning (McNeill, 1992; Kendon, 2004; Poggi, 2007). Gestures can also be studied according to their functions in the communication process. For example, they can have a demarcative function and mark the rhythm of an utterance, so as to underline speech chunks or to coordinate who has the speaking turn. They can also be tightly tied to dialog acts underlying a speaker's intention. But gestures can also reveal a speaker's attitude toward what she is saying such as her level of certainty or of agreement. Additionally, gestures can carry information about affective states (Bänziger et al., 2012).

To convey these varieties of functions, the form and timing of gesture production in relation to speech is critical. The temporal relationship between speech and gesture is far from being trivial as gesture can coincide with speech prosody or can be anticipated or maintained afterward (Kendon, 2004). Additionally, gesture shape and movement carry important meaning.

In Wagner et al. (2014), the authors gave an extensive review of work on communicative gestures, from psychology studies to computer systems. The results highlighted how closely tied together speech and gesture are (in terms of meaning and timing). According to some theoretical models, like McNeill's *Growth Point Theory* (McNeill, 1985), this could be explained by the fact that gestures and speech are produced from the same mental process. In particular, many studies investigated the effect of embodied cognition on speech and gesture production (Hostetter and Alibali, 2008) and hypothesized the existence of a common mental imagery between the two communicative channels (Kendon, 1980).

2.1. Gestures—Types and Structures

Some scholars have underlined how gesture definitions, in term of shape and movement, can be viewed as the abstraction of an

action (Kendon, 1980; Calbris, 2011). This is particularly true for metaphoric gestures. For example, rejecting an idea can be conveyed by a hand gesture metaphorically mimicking rejection with a pushing away gesture.

Gesture can be characterized by its physical constituents. The form of a gesture is described in term of the shape of the hand, the wrist and the palm orientation. A gesture can be made with one or two hands, symmetrically or in opposition. The movement of a gesture can be defined by its direction, its path, its dynamism.

As mentioned by Kendon (1980), gestures exhibit different structures. At the level of a gesture, there are different phases (e.g., preparation, stroke, hold and relaxation). Consecutive gestures can be co-articulated, meaning that the last phase of a gesture is mapped to the beginning phase of the next gesture. There is a higher structure that corresponds to discourse segments in which consecutive gestures share some of their constituents and are kinetically segmented. It corresponds to the ideational structure introduced by Calbris (2011). In her theory, Calbris argues that discourse is composed of units of meaning and rhythm she calls *Ideational Units*. Within an Ideational Unit, there is a consistency between the gestures of the person as they show similar properties.

2.2. Conceptual Metaphors and Image Schemas

Within the literature on embodied cognition, the *conceptualization hypothesis* states that the way we mentally represent our world is constrained by our body (Wilson and Golonka, 2013). In other words, our interactions with the world through embodiment lead to the conceptual representations we manipulate in our mind to ground abstract and concrete concepts. This is how we can apply physical properties to abstract concepts as part of our metaphorical reasoning. Lakoff and Johnson (1980) describe *Conceptual Metaphors* to explain how we can talk about one domain using properties from another one. For instance, in the conceptual metaphor LOVE IS A JOURNEY, love is seen as having an origin, a destination (might be an end) and a series of events or steps between the two.

In that case, how do we represent in our mind these properties that can be shared between concrete and abstract entities? Johnson suggested that humans use recurring patterns of reasoning, called *Image Schemas*, to map these conceptual metaphors from an entity to another (Johnson, 1987). These *Image Schemas* have also been studied by Grady in order to attempt to explain how our perception mechanisms are at the origin of our metaphorical reasoning (Grady, 2005).

For instance, the *Image Schema* CONTAINER gives an entity the typical properties of a container such as having a boundary with elements that are within it and elements that are outside. We can think of culture metaphorically as a container in terms of people that are part of the culture, and people that are not. This illustrates how people use their physical reality to reason about abstract concepts, thus giving physical attributes to abstract concepts. Moreover, according to Wilson, using metaphoric reasoning can unconsciously influence our nonverbal behavior: if someone is thinking about a future event, he might be swaying slightly forward (Wilson and Golonka, 2013).

2.3. Image Schemas and Gestures

While these *Image Schemas* have been investigated as linguistic structures (Croft and Cruse, 2004), used in the production of speech, other work suggests that they could be at the origin of the accompanying gesture production as well (Cienki, 2013). In Mittelberg (2008), the author describes how a gesture (mimicking the shape of a box) can represent the *Image Schema* OBJECT or CONTAINER, itself being linked to the conceptual metaphor IDEAS are OBJECTS. In other work, Cienki conducted an experiment to study if *Image Schemas* (a subset) could be used to characterize gestures (Cienki, 2005); his conclusions showed positive results. In Chui (2011), the authors revealed evidence of the use of spatial conceptual metaphors in gesture production for mandarin speakers. Another experiment by Lücking and his colleagues tried to find recurrent gestures features in the expression of particular *Image Schemas* (Lücking et al., 2016). Their results showed that, for some *Image Schemas*, people spontaneously used similar gesture features. Finally, in Mehler et al. (2015), the authors developed a gesture-based interface for an interactive museum system that used *Image Schemas* as a basis for their gestural grammar.

Metaphorical reasoning allows the transfer of properties from a source domain to a target domain and, in the discourse, this is realized by talking about the target domain as if it was an entity of the source domain (Lakoff and Johnson, 1980). Metaphoric gestures follow a similar process and, like iconic gestures, their characteristics serve to illustrate and demonstrate particular physical properties (metaphorically projected in the case of metaphoric gestures) of the concept being communicated by the speaker (Cienki, 1998). A hypothesis is that these characteristics are tied to the *Image Schemas* underlying the production of the metaphorical reasoning. Researchers have highlighted how some typical metaphorical properties are often represented with the same gesture characteristics (Cienki, 1998; McNeill, 1992; Calbris, 2011). For instance, to represent the CONTAINER concept, one might exhibit concave hands facing each other in a bowl-like shape. These findings are in line with earlier works of McNeill and Levy who observed how people (through the use of the conduit metaphor Reddy, 1979) illustrate an abstract entity, which could be tied to the OBJECT *Image Schema*, by pretending to hold an object with their hand (McNeill and Levy, 1980).

More examples are given in these works. They illustrate that different characteristics are used depending on the metaphorical element being portrayed. Whereas CONTAINER and OBJECT seem to be depicted through the shape and the orientation of the hand, other *Image Schemas* can be portrayed by other physical characteristics such as the position or the quality of the movement. The *Image Schema* SPLIT, which would underlie a separation or a difference, can be illustrated by a vertically flat hand moving abruptly downward; the SCALE *Image Schema*, parameterized so it encapsulates the action of an increasing scale, can be depicted with both hands moving away from each other (Calbris, 2011).

Inspired by this research, we propose to use *Image Schemas* as the basis for our representation, to bridge the speech of an ECA and its gestures.

2.4. Gestures and Speech Alignment

Timing is key to the alignment of speech and gestures. For example, in McNeill's Growth Point Theory McNeill (1985), the growth point is the initial form (or seed) of the thinking process from which the future speech and gesture are constructed together in synchrony with each other. While *Image Schemas* are good candidates for predicting gesture shapes, additional information will be required in order to identify the most appropriate meaning to be aligned with the speech by the gesture production (not all *Image Schemas* are turned into gesture; selection happens). Even if each word in isolation carried an embodied meaning represented by an *Image Schema*, people do not produce a gesture on every word. For instance, an "important obstacle" potentially represents two *Image Schemas*, SCALE (parameterized to encapsulate a big quantity) and BLOCKAGE. However a speaker might produce a single gesture (i.e., overlapping the pronunciation of both words) corresponding to the meaning that is being emphasized in the context of the conversation.

Prosodic and linguistic features of the speech seem to have the potential to be the contextual markers that could be correlated with the *Image Schema* selection process (Wagner et al., 2014). Several works showed that gesture and speech timings seem to be close to each other but not exactly simultaneous. Results from Leonard and Cummins (2011) or Loehr (2012) acknowledge the correlation between gesture phases and prosodic markers while accepting slight variations. In the particular case of beat gestures, which are not constrained by meaning, the peak of the stroke seemed to be closer to the pitch emphasis (Terkén, 1991). For representational gestures, it would seem that the gesture anticipates the prosodic markers of the discourse. In Kendon (1980), Kendon states that the stroke of a gesture precedes or ends at, but does not follow, the phonological peak of the utterance. In her work, Calbris also demonstrated that when constructing thoughts in a discourse, gestures tend to slightly anticipate the speech (Calbris, 2011).

Additionally, an utterance can be decomposed into a theme, the topic being discussed, and a rheme, the new information on the theme that is being conveyed (Halliday et al., 2014). Calbris observed that while enunciating the rheme of an utterance, more representational gestures are produced than in the theme (where more beat and incomplete gestures are produced) (Calbris, 2011). In other words, people tend to produce more representational gestures for accompanying and describing the new information brought by the rheme, and would align the peak of the gestures so it falls closely in time with the accentuation of the pronunciation.

3. PRODUCING COMPUTATIONALLY COMMUNICATIVE GESTURES

Different approaches have been investigated to address the challenge of automating gesture production and more precisely communicative gestures. Much of the existing work proposes independent reasoning units that dissociate gesture production from speech production. For instance, in Thiebaut et al. (2008), the authors developed a Behavior Realizer (Vilhjálmsdóttir et al.,

2007) capable of using different kinds of animations (computed in real-time or using pre-configured handcrafted or motion captured animations) to perform a set of requested signals. Their architecture is structured into different components communicating through a messaging system, allowing for a dynamic and responsive system and they introduced hierarchical rules to blend lower bodily functions (like posture) with higher level ones (like gaze). In the following, we present other studies that tried to do gesture alignment with the prosody or direct mapping from the surface text of the agent's discourse to gestures.

In Levine et al. (2009), the authors develop a real time system that produces gestures using prosody as input and Hidden Markov Models as the probabilistic gesture model. This model was not capable of properly handling the alignment between prosodic cues and gesture segments so in Levine et al. (2010), the authors proposed an improved version of the model using Conditional Random Fields. The result is interesting as their system produces well-aligned gestures but their meaning (and therefore the gesture shape) is not correlated with the content of the speech.

In an effort to produce gestures that were both well-aligned as well as correlated with the speech content, Chiu and Marsella integrated several data-driven, machine learning approaches¹ to acquire a model that took lexical, syntactic and prosodic features as input (Chiu and Marsella, 2014; Chiu et al., 2015). While the approach was capable of producing well-aligned gestures correlated with the content, the results also illustrated that using machine learning to realize automatic gesture production capable of the richness of human gesture production would require a far more extensive data collection effort.

Lee and Marsella (2006) compare two approaches to generate nonverbal behaviors. The first approach, called the *literature based* approach, involves using the literature on nonverbal behavior as well as manual analysis of videos of human-human interaction to hand craft rules that map between the content of human speech and gestures. The overall design effort and complexity of such rule-based systems is very high. The second approach, a *machine learning* approach, uses a data-driven automated processes to find features (in the AMI meeting corpus Carletta et al., 2006) that are strongly associated with particular behaviors. Then, one can use those features to train models that will predict the occurrences of the behavior. The authors compare several different learning techniques (Hidden Markov Models, Conditional Random Fields, Latent-Dynamic Conditional Random Field) on syntactic features, dialogue acts and paralinguistic features, to predict speaker's head nods and eyebrow movements. The same authors used a machine learning approach in Lee and Marsella (2009) to automatically produce head movements on each part of the speech according to the dialog acts and the affective state of the agent.

Sargin et al. (2008) developed a two-level Hidden Markov Model for prosody driven head-gesture animation where the first

¹The work combined deep learning techniques with the temporal modeling capabilities of Conditional Random Fields to select which gesture to convey the meaning, while Gaussian Process Latent Variable Models were used to synthesize the gesture motion and co-articulate the gesture sequences.

level performs temporal clustering while the second layer does the joint modeling of prosody-gesture patterns.

In Busso et al. (2005), the authors synthesize rigid head motion from prosodic features, they also perform canonical correlation analysis to ascertain the relationship between head motions and acoustic prosodic features. The results suggested that head motions produced by people during normal speech are very different from the motions produced with an emotional state.

While prosodic information has been shown to be relevant to identify the timing and the intensity of gestures, making it a powerful input for generating beat gestures with no particular meaning or connection to the verbal content, producing representational gestures (deictic, iconic or metaphoric) requires an understanding of the information that the speaker wants to convey. Researchers aiming at producing automatically representational gestures synchronized with speech have looked at the potential of using the surface text of speech to link it with gestural representation that can convey similar or complementary information.

In Bergmann and Kopp (2009), the authors learned from an annotated corpus of spatial descriptions a Bayesian model used to predict the shape of a speaker's iconic gestures to describe the shapes of objects situated in a virtual environment (like a church). The shape of the iconic gestures is automatically computed from a geometric description of the objects in the environment. Such an approach was also used in Kopp et al. (2007) where the authors established *Image Descriptive Features* IDF (conceptually close to *Image Schemas* but used to describe geometrical and spatial features of concrete entities) and how they relate to gesture features. In both works (Kopp et al., 2007; Bergmann and Kopp, 2009), their context was a direction-giving task. They analyzed a corpus of interaction between person giving directions and exposed evidences of correspondence between the gesture features and the spatial features of the object being described. While both system allow combining multiple IDFs or geometrical description to form one gesture, which is the approach we are considering, they do not take into account the transfer of gesture properties throughout the utterance of the agent.

In Kipp et al. (2007), the authors detail their data-driven approach to build a system able to automatically generate gestures synchronized with speech. Their approach relies on the annotation of a corpus of videos of a speaker, identifying her gestures and the words associated with them, which is then used to learn the probabilities to observe particular gestures with particular words (reduced to semantic tags). Their system is capable of handling the co-articulation of gestures. When generating and selecting gestures, proximity among gestures (in terms of timing) is used to group them into gesture phrases. This grouping allows for the adaptation of the different phase existences and timings to co-articulate gestures within the same phrase and is realized thanks to a set of rules and constraints.

In most of the reviewed works, the proposed systems either tackle one aspect of our challenges (the alignment or the semantic depiction) or do not consider an intermediate representation that would allow them to reason on agent's mental state and to extend the gesture production with additional communicative intentions (such as the expression of emotion). The work that

is the closest to our approach is the work conducted by Marsella and his colleagues to develop the Cerebella system (Marsella et al., 2013; Lhommet et al., 2015). In Cerebella, the studies of Lhommet (Lhommet and Marsella, 2014, 2016) and Xu (Xu et al., 2014) were combined into a complete system that extracts a mental representation from the communicative intentions of the agent to produce corresponding gestures.

In Lhommet and Marsella (2016), the authors proposed a model that maps the communicative intentions of an agent to primary metaphors in order to build a mental state of *Image Schemas*. This mental state is used to produce corresponding gestures in a second stage. In Xu et al. (2014), the authors propose a system that produces sequences of gestures that respect the notion of Ideational Units. Their system accepts as input communicative functions organized within Ideational Units (using an augmented version of the Functional Markup Language Heylen et al., 2008). This information is used to generate, using a set of defined constraints and rules, gestures that share some properties (ex. shape of the hand or location) or are co-articulated when belonging to the same Ideational Unit. However, in this work, the authors limited themselves to a restricted subset of *Image Schemas* and therefore have a limited potential for generalization.

In our work, we aim at proposing an architecture for automatically computing communicative gestures inspired by the different aspects of the challenges that have been investigated by previous researchers. Our model takes into account the linguistic structure, the prosodic information and a representation of the meaning conveyed by the agent's speech to derive gesture characteristics that are combined into coherent gesture phrases thanks to an Ideational Unit mechanism. Our model is geared to integrate a richer representation of *Image Schemas* and to be integrated in an agent system that computes in real-time the multimodal behaviors linked to additional communicative functions (such as showing emotional states and attitudes).

4. IMAGE SCHEMA BASED GESTURE GENERATOR

If we were trying to replicate cognitive models proposed in the literature (e.g., Barsalou, 2009), We would need to represent mental states, and additional components such as the agent's perception, to build its inner reasoning pattern (Wilson and Golonka, 2013). As a first step, we prefer to adopt a simplified approach where *Image Schemas* are immediately tied to the speech and the gestures. We make this assumption for following reasons.

First, the focus of our investigation is on the meaning conveyed by both the verbal and the nonverbal channels. Therefore, we particularly stress the importance of the mental imagery we chose to fulfill this task. We do not reject the idea that a more faithful model would need to integrate additional reasoning components such as a grounding mechanism like in Lhommet and Marsella (2016). But our efforts focus on identifying if using a shared language between speech and

gestures allows for generating more consistent multi-modal behaviors.

Second, we would have to perform more investigation on how to replicate embodied cognition mechanisms within the virtual environment of the ECA. Embodied cognition is related to the physicality of our experiences and a virtual agent does not physically experience its environment (even if it could be simulated). This is a very interesting line of research but its perspectives are outside the scope of our objectives.

The model we propose is organized around the concept of *Image Schemas* as the intermediate language between the verbal and nonverbal channels. We propose an adaptation of the theoretical framework shown in **Figure 1**. In order to be compatible with existing speech production system, our system takes as input the speech of the agent with the prosodic markers, infers possible *Image Schemas* underlying the speech and generates the corresponding gestures. In the future, we might have a speech production system that works with *Image Schemas* and therefore which is capable of giving these *Image Schemas* to the gesture production component. But for now, we have to find a way to extract them from the text. Our architecture is composed of three levels: an *Image Schema* extractor, a gesture modeler and a behavior realizer supporting *Ideational Units*. This architecture is shown **Figure 3**.

4.1. Image Schema Extractor

The *Image Schemas* extraction component has the task of identifying the *Image Schemas* from the surface text of the agent's speech and to align them properly with the spoken utterance (for future gesture alignment). However, there does not exist a definitive list of *Image Schemas* and different researchers have proposed complementary or alternative ones. Therefore, we propose our own list adapted from the original list of Johnson (1987) and Clausner and Croft (1999). Following the idea of a parameterization of *Image Schemas* (each *Image Schema* could

have different values), we decompose the SCALE *Image Schema* into smaller ordered units that would be more easily exploitable at a computer level (SMALL, BIG, GROWING, REDUCING) resulting in the following list: UP, DOWN, FRONT, BACK, LEFT, RIGHT, NEAR, FAR, INTERVAL, BIG, SMALL, GROWING, REDUCING, CONTAINER, IN, OUT, SURFACE, FULL, EMPTY, ENABLEMENT, ATTRACTION, SPLIT, WHOLE, LINK, OBJECT. This list allows us to manipulate spatial, temporal and compositional concepts (container vs. object and whole vs. split for instance). This list is not exhaustive and should definitely evolve in the future. This does not only mean adding new *Image Schemas*, but also enriching their representation. It should be possible later to parameterize *Image Schemas*, so that the gestures can be parameterized as well, and to combine them together. For instance, it should be possible to connect *Image Schemas* together to describe the evolution of an entity being discussed, like a CONTAINER being FULL or an OBJECT being an ATTRACTION or part of a SPLIT. For now, we are adopting a simplification where we are only looking to find an unparameterized *Image Schema* to match it with a gesture invariant. Gesture invariant corresponds to a feature of a gesture that is always present to carry a given meaning Calbris (2011). Our assumption is that as a first step, producing the invariant should result in a coherent animation in terms of meaning.

4.2. Gesture Modeler

After obtaining a list of aligned *Image Schemas* for a sequence of spoken text, the gesture modeler builds the corresponding gestures.

The first step is to retrieve the gesture invariants to build the final gestures. According to the literature, the typical features of a gesture are: hand shape, orientation, movement and position in gesture space (Bressem, 2013). In Kopp et al. (2007), the authors proposed to represent gestures using the first three features augmented with a movement information on each of them. In our work, for each *Image Schemas* we want to find which features are needed to express its meaning and how it is expressed. For this task, we propose a dictionary that maps each *Image Schema* to its corresponding invariants (the features that need not to be altered to properly express the meaning). This dictionary is depicted in **Table 1**. This dictionary was conceived after a review of work on gesture meaning (Kendon, 2004; Calbris, 2011) and contains the minimal features required to express a specific *Image Schema*. It is not fixed and can be expanded.

Once the invariants are retrieved, a gesture is built using two default gesture phases (a beginning and an end) parameterized to reflect the specific invariants. Since we are using a default template for the phases, most of the motion is predetermined but the use of the specific invariants alters significantly the shape of the gesture to express the desired meaning. For instance, if a gesture should encapsulate the *Image Schema* UP, a gesture will be built with its second phase (the stroke) that goes through a high position. In order to decide what a high position is, we follow McNeill's gesture space that divides the space used by the hand while gesticulating into 18 subspaces (upper position, lower position, periphery, center etc.) (McNeill, 1992).

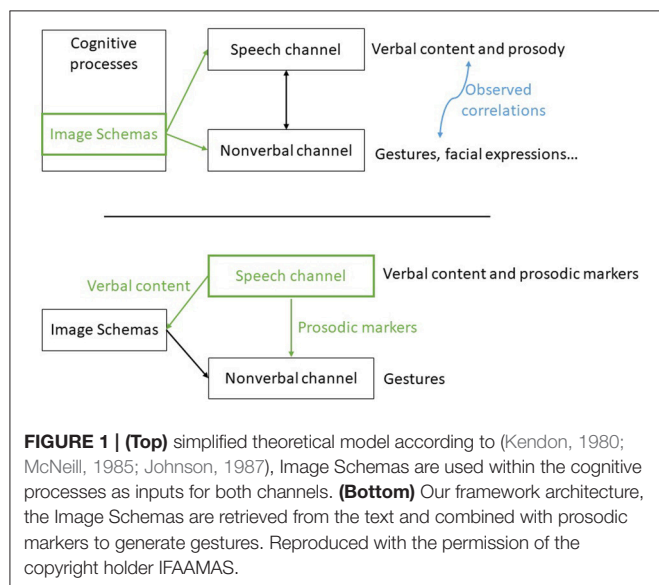


TABLE 1 | Association between image schemas and invariant gesture characteristics.

Image schemas	Handshape	Position	Orientation	Movement
UP		Up		
DOWN		Down		
FRONT		Front		
BACK		Close	Back	
LEFT		Left		
RIGHT		Right		
NEAR		Close Center		
FAR		Away	Frontward/ Downward	
INTERVAL	Flat		Inward	
BIG	Open	Away	Inward	
SMALL	Mid-closed	Close center	Inward	
GROWING				From SMALL to BIG
REDUCING				From BIG to SMALL
CONTAINER	Bowl-shape		Inward	
IN	Picking-shape		Downward	
OUT	Open spread		Outward	
SURFACE	Flat		Downward	Horizontal wipe
FULL	Closed fist			
EMPTY	Open spread			
ENABLEMENT	Open			Frontward
ATTRACTION	Closed fist			Backward
SPLIT	Flat		Inward	Abrupt downward
WHOLE	Open		Inward	
LINK	Hold			Translation
OBJECT	Conduit shape			

4.3. Behavior Realizer Using Ideational Units

The final layer of our framework has the role of combining the composed gesture obtained through the previous components to produce the final animation of the virtual agent.

We define a system that follows the *Ideational Unit* model proposed by Calbris (2011) and the computational model of Xu et al. (2014). The system operates the following main functions: (1) co-articulates gestures within an *Ideational Unit* by computing either a hold or an intermediate relaxed pose between successive gestures (instead of returning to a rest pose), (2) transfers properties of the main gesture onto the variant properties of the other gestures of the same *Ideational Unit*, (3) ensures that a meaning expressed through an invariant is carried on the same hand throughout an *Ideational Unit* and (4) finally dynamically raises the speed and amplitude of repeated gestures. More precisely, to compute the relax pose of a gesture, our algorithm lowers the wrist position in 3D space; it also modifies the hand shape by using the relax position of the fingers rather

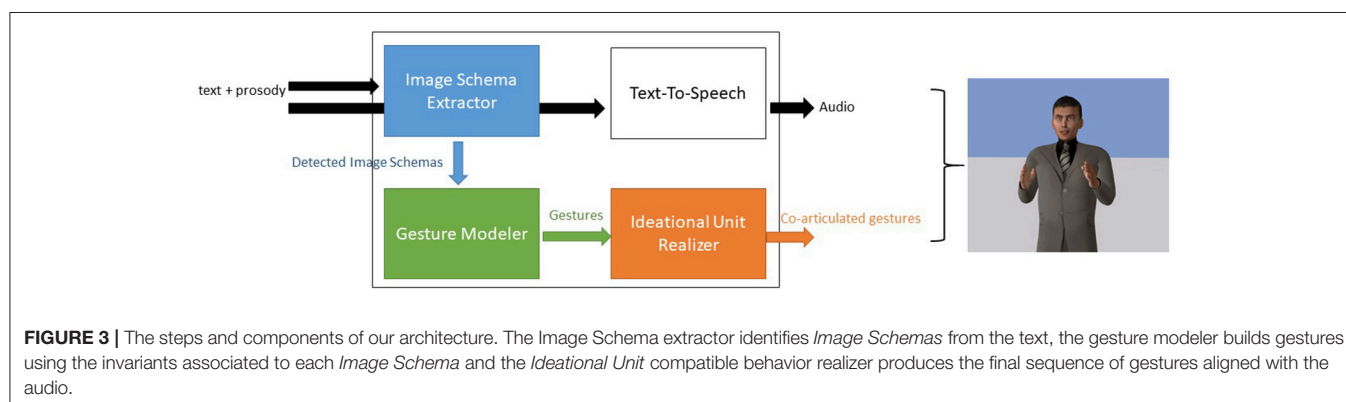
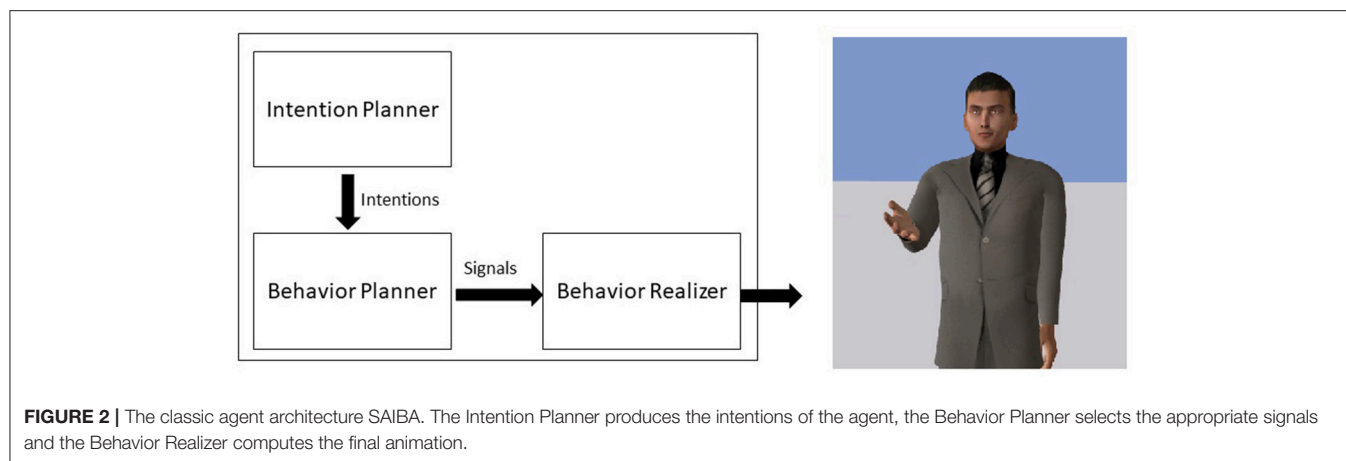
than straight or closed positions. A gesture phase is held within an *Ideational Unit* when the time between the end of the gesture stroke and the beginning of the next gesture stroke is below a given threshold. To transfer properties of one gesture (here the main gesture) to the other ones, we configure their features to be identical to the main gesture, unless they were indicated as invariant. To mark the repetition of a gesture, we extend the position of the wrist in 3D space for each gesture stroke position to increase the amplitude of the gesture. We do not modify the timing of the gesture phases but since the position of the arms have been extended and their duration is the same, the speed is increased as a consequence.

This mechanism needs to know which is the main gesture of an *Ideational Unit* and what are the invariants of the gestures (in order to know which features from the main gesture can be copied to which features of the other gestures). This information is found within our dictionary of invariants. We are not working on the automatic detection of *Ideational Unit* in the text however, since this information is needed, we proposed a simplification of the approach that considers for now that a sentence is equivalent to an *Ideational Unit*. Of course, an *Ideational Unit* can span over multiple sentences or multiple *Ideational Units* could be found in a sentence, but this approximation allows us to start to manipulate this concept. In order to select the main gesture, we follow this simple rule inspired by Calbris' observations on the importance of the rhyme in a sentence Calbris (2011): we choose as the main gesture the first gesture, in the sentence, built from a stressed *Image Schema* (using the prosodic markers).

4.4. First Implementation : Metaphoric Gesture Generation

In order to assess the relevance of our approach, we implemented a preliminary version of the system that focuses on the production of metaphoric gestures found in political speeches. We decided to explore political speeches since they are known to be richer in conceptual metaphors (Lakoff and Johnson, 1980), which in turn might lead to more metaphoric gestures (Cienki, 1998) that, according to our assumptions, should convey *Image Schemas*.

We implemented the model within the agent platform Greta (Pecune et al., 2014). Greta is an agent platform, compliant with the SAIBA standard, that allows the development of components that integrate seamlessly. The SAIBA standard defines the base components of an agent which includes an Intention Planner, in charge of computing the communicative intentions of the agent, a Behavior Planner, in charge of selecting the different signals (verbal and nonverbal) to perform the intentions and a Behavior Realizer that produces the final animations (see Figure 2). In our case, we developed the *Image Schema* Extractor and the Gesture Modeler as an alternative to the Intention and Behavior Planners and we extended the Behavior Realizer in order to take into account *Ideational Units* in the production of the animations (see Figure 3). The system reads an XML-based text file (a Behavior Markup Language BML document as described in Vilhjálmsdóttir et al., 2007) that describes the textual speech of the agent marked with prosodic and *Ideational Unit* information and produces



the complete animation with the audio using a Text-To-Speech component.

4.4.1. Image Schemas Extractor

Algorithm 1 ImageSchema extraction using WordNet.

```

for all word do
  ImageSchema = none;
  SynonymSet = Lesk(word);
  while TopNotReached() & ImageSchema == none do
    ImageSchema = getImageSchema(SynonymSet);
    MoveUpFollowingHypernym();
  end while
end for

```

For this first implementation of the *Image Schema* extractor, we are using an expert approach using the WordNet dictionary (Miller, 1995). In WordNet, words are organized in synonym sets. A synonym set represents a meaning, with all the words belonging to a synonym set sharing the same meaning. Each set is connected to other sets by semantic relations, giving additional information on a particular meaning. Following the hypernymic relations of a synonym set, one can obtain a synonym set with a more general meaning (for instance a hypernym of *table* is

furniture). This organization is similar to a class inheritance system.

It is important to mention that a word might belong to different synonym sets if it can have multiple meanings. For instance, the word *table* can mean a piece of furniture or a set of data arranged in rows and columns.

Our algorithm works as follows (see Algorithm 1): for each word in the text, we use the Lesk method to disambiguate the meaning of the word and find the most likely synonym set for it using WordNet (Lesk, 1986). The Lesk algorithm compares the set of neighbors of the word being analyzed, in the current sentence, with its different definitions and chooses the definition (the synonym set) that has the most words in common with the neighbors. Then, we follow the hypernym path up in the hierarchy until we find a synonym set corresponding to our *Image Schemas* (if none is found, no *Image Schema* is returned). Using the literature on conceptual metaphors and by observing political videos, we empirically established this repertoire of synonym sets corresponding to *Image Schemas*. Several synonym sets are associated to each *Image Schema* to cover possible variations in meaning.

4.4.2. Syntactic and Prosodic Selection

Instead of keeping all *Image Schemas* that were detected for every word, we select some of them by following observations from the

literature in order to avoid exaggerating the gesticulations of the agent. We use OpenNLP chunker (Morton et al., 2005) to group words into phrases (e.g., noun phrases and verb phrases) and we tag one *Image Schema* per group as the main *Image Schema* of this group. We use the Stanford POS Tagger (Toutanova et al., 2003) to retrieve the syntactic role of each word and we prioritize the *Image Schemas* obtained from modifiers such as adverbs and adjectives (Calbris, 2011) as main ones unless a stressed accent is put on a particular word, in which case we prioritize the *Image Schema* coming from this word. This also leads to the selection of the main gesture of an *Ideational Unit* as seen in section 4.3. In case of multiple candidates, we randomly select the *Image Schema* for the group from them. As we saw earlier in section 2, gestures can also slightly anticipate speech (Wagner et al., 2014). In order to properly align them, we use the prosodic information to ensure that gesture strokes end at or before (up to 200 ms) pitch accents (Kendon, 2004). In Wang and Neff (2013), the authors identified through an experiment that an agent's gestures might not need to be tightly synchronized, little variations are acceptable, but should they arise, gestures should be moved earlier and not later (which is comparable to what has been found in the literature). The result is a list of *Image Schemas*, each one specifying exactly when it starts and ends in the spoken text using time markers. The prosodic information needs to be given to our system. We developed a pipeline to transform videos with subtitles into our BML format that describes the speech content (as text) along with its pitch contour. We are using OpenSmile (Eyben et al., 2013) to extract the pitch contour and *gentle*² speech alignment tool to align the words with it. From there we can automatically build the BML files, that include the prosodic information associated with the text content of the speech, ready to be given to our system to generate a corpus of examples.

4.4.3. Illustration

To illustrate our gesture generator model, we selected a video³ showing a politician (Al Gore) displaying metaphoric gestures; we transcribed the textual speech and the prosodic information from the videos and let our system produce the corresponding gestures⁴. In this video, Al Gore is producing many metaphoric gestures. This video offers an interesting comparative basis to see if our model can capture the invariants of these metaphoric gestures. The output of our gesture generator model showed similarity with the input video. For each metaphoric gesture of the video, our model produced a gesture with similar timings. Some of them were carrying similar meaning as well; for the sentence “the internet is full of junk,” both the politician and our system produced a circling gesture depicting the fullness underlined in this sentence. In another example, at the beginning, the politician says “we have to get back to harvesting the wisdom of crowds” while moving his arms in a circle like he is gathering the wisdom (see Figure 4). Our algorithm captured the *Image Schema* ATTRACTION in the word *harvesting* and therefore, produced a gesture where the agent is pulling something toward her (very similar to the politician gesture).

Another interesting example happened when the politician said “good ideas rise to the surface.” In the video, the politician does a gesture mimicking something going up, to accompany the verb “rise.” In our output, the *Image Schema* SURFACE, extracted from the word “surface,” was identified as the main *Image Schema* of the group rather than the UP one (that was extracted for the “rise” word). This choice resulted in the agent doing a gesture with a horizontal wipe (see Figure 5). This example is interesting as, despite being different in meaning (compared to the politician original gesture), the gesture produced by our system was still coherent with the words of the speech. In the original video, the temporal relationship between the speech and gestures varies, with gestures being perfectly in sync and others being a little bit ahead of the speech, consistent with the literature on the timing of gestures. Our system did not produce that much variability in the temporal relationship between speech and gesture, resulting in gestures having closer temporal relationship with speech in our output than in the original video. Understanding what causes this temporal variability in human communication in order to model it is another challenge that could be addressed in future work.

We observed that the output of our system did not systematically reproduce the exact gestures seen in the source video as it may select other *Image Schemas* to be highlighted with a gesture (linked to another intention the speaker wants to convey); but, nevertheless, it was able to generate animated sequences that are coherent in terms of speech-gesture mapping and synchronization. We gave as input to our gesture generator model Al Gore's speech defined in term of words and prosody. Such input does not capture all the speaker's intentions. Differences between Al Gore's gestures and output from our algorithm could arise from this lack of information. Our algorithm uses only words and acoustic information to select which metaphoric gesture to display. It does not catch which intention prevails in selecting a gesture. In the example “good ideas rise to the surface” Al Gore does a UP gesture emphasizing the emergence of good ideas while our model computed a SURFACE gesture as specified in the text.

In a traditional SAIBA architecture, we start from the intention of the agent from which we derive the signals to produce. In our system, we assume that the speech is given to us, without describing exactly what was the original intention that led to this speech. This information would be useful in order to disambiguate the meanings and to identify which word should be stressed and illustrated with a gesture. Arguably, looking to retrieve the *Image Schema* is a first step toward a mechanism that could retrieve the communicative intentions of a speaker but this is out of the scope of the current work.

5. CONCLUSION

Throughout this article, we established the foundations for developing systems capable of generating metaphoric gestures automatically from speech. We identified the key challenges for the completion of this objective, from the synchronization of speech and gestures in terms of rhythm and intensity, to a proper meaning representation and the conveying of that meaning. We discussed some of the fundamental issues raised in the psychological and embodied cognition literature

²<https://lowerquality.com/gentle/>

³<https://youtu.be/0ggic7bDNSE>

⁴<https://youtu.be/47QLONZS5zw>

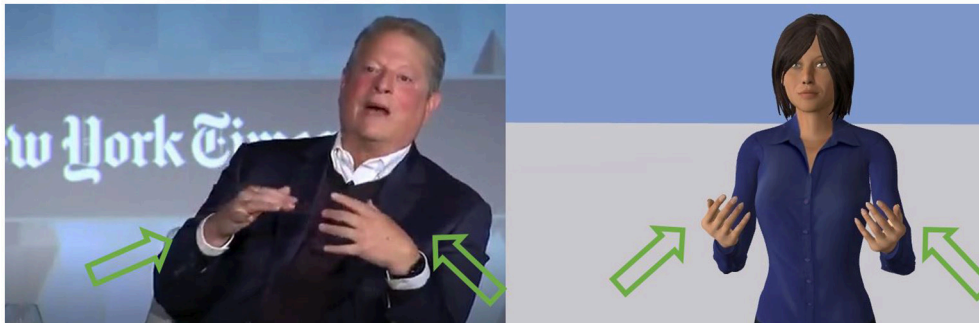


FIGURE 4 | The case of the sentence “We have to get back to harvesting the wisdom of crowds.” **(Left)** The politician gather the space around him toward him. **(Right)** The agent produces a similar gesture by pulling something toward her. Reproduced with the permission of the copyright holder IFAAMAS.

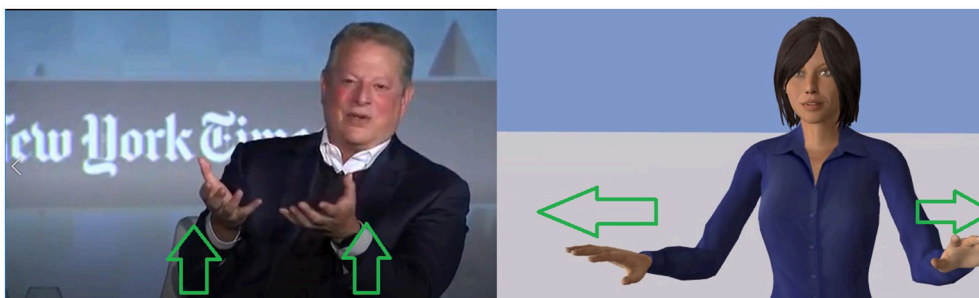


FIGURE 5 | The case of the sentence “good ideas rise to the surface.” **(Left)** The politician illustrates his speech with a rising gesture, communicating a particular intention. **(Right)** The agent choses to illustrate the surface concept and thus displays an horizontal wipe gesture which illustrate a different communicative intention. Reproduced with the permission of the copyright holder IFAAMAS.

on how people build and use structured representations to produce both verbal and nonverbal behaviors. From there, we proposed to use an intermediate representation between the text and the gestures inspired by *Image Schemas* that could help us solve the technical challenges of computing automatically the communicative gestures. Our approach relies on inferring automatically from the surface text of the agent the possible underlying *Image Schemas* and to combine those with the prosodic information in order to select the particular gesture characteristics to convey the imagery. In order to propose a coherent and flexible system, this process is integrated with an ideational unit compatible engine that takes care of invariant priority and co-articulation between the gestures. Our approach leverages previous studies that tackled various parts of our objectives by extending some of their functionalities and by combining them into one complete system with regards to the existing agent's standards (the SAIBA architecture). These parts include how to synchronize gestures and speech based on prosodic information, how to configure the characteristics of the gestures (hand shape, movement, orientation) to convey the desired representational meaning and finally how to combine and co-articulate these gestures into a coherent and meaningful unit of behaviors. We implemented a first version of the system in order to evaluate the potential of our approach. Our method does not always produce the same gestures as in an original

video. From a technological perspective, these differences mainly come from the selection of the “important” *Image Schema* and with the speech alignment. A potential improvement for this approach would be to use Sequential Learning approach as they have proven to be an effective method to identify particular structures in text like Named Entities (Nadeau and Sekine, 2007). Additionally, we consider an utterance and its prosody profile, but we do not take into consideration other contextual factors such as what has already been said or if there are contrastive elements in the utterance. Another explanation for the differences we obtained could be that our system has a limited set of gesture invariants and, despite being able to produce coherent gestures, it cannot capture the variations or style of a speaker. An interesting alternative could be to build a stochastic model of invariants learned from a corpus of gesture data for a given speaker. This could introduce more variability and allow the reproduction of a “speaker style” like in Durupinar et al. (2016).

Alternatively, these differences might be due to some other limitations from the theoretical background we are relying on. The idea of a common mental structure is quite developed through the literature as seen by work such as Croft and Cruse (2004) or Cienki (2013) but the exact mechanism is still unknown. While the field of embodied cognition supports the idea that our physical interactions with the world shapes

these structures (see Johnson, 1987; Wilson and Golonka, 2013), the process that could give different shapes (how big is big in the speaker's mind?) and importances to them (what does the speaker want to emphasize?) remains a complex system which is not fully understood yet. Exploring these models and theories with the use of virtual character capable of mimicking the human communication processes, which can be extended and manipulated, could help to investigate the details of these theories.

Whereas our approach will allow an agent to produce automatically metaphoric gestures, more investigation has to be done to ensure how to extend our system to handle other representational gestures (like deictic and iconic). Moreover, a challenge that will arise will be to combine the meaning conveyed by these metaphorical representations with the communicative intentions of the agent or with other nonverbal behaviors that can be used for turn regulation in the conversation. In the near future, we plan to evaluate our model through a perception study where participants will assess which *Image Schema* they perceive in the gestures of the virtual agent. Their feedback will be valuable

to assess the progress of our approach toward an automatic generation of nonverbal behaviors as well as to inform the next steps of our research.

AUTHOR CONTRIBUTIONS

BR wrote the manuscript with support and feedback from all the other authors. More specifically, CC participated in the review of work related to the use of prosodic information for gesture alignment. CP contributed to the formal definition of gestures used in the manuscript. SM provided assistance with the theoretical definitions coming from the embodied cognition domain.

FUNDING

This work was supported by European projects H2020 ANIMATAS (ITN 7659552) and H2020 Council of Coaches (SCI 769553).

REFERENCES

- Argyle, M. (1972). "Non-verbal communication in human social interaction," in *Non-Verbal Communication*, ed R. A. Hinde (Oxford, UK: Cambridge University Press).
- Argyle, M. (1972). Non-verbal communication in human social interaction. In R. A. Hinde, Non-verbal communication. Oxford, England: Cambridge U. Press.
- Bänziger, T., Mortillaro, M., and Scherer, K. R. (2012). Introducing the geneva multimodal expression corpus for experimental research on emotion perception. *Emotion* 12:1161. doi: 10.1037/a0025827
- Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behav. Brain Sci.* 22, 637–660. doi: 10.1017/S0140525X99532147
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 1281–1289. doi: 10.1098/rstb.2008.0319
- Bergmann, K., and Kopp, S. (2009). "Increasing the expressiveness of virtual agents: autonomous generation of speech and gesture for spatial description tasks," in *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1* (Budapest: International Foundation for Autonomous Agents and Multiagent Systems), 361–368.
- Bressem, J. (2013). "A linguistic perspective on the notation of form features in gestures," in *Body–Language–Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 1, eds C. Muller, A. Cienki, E. Fricke, H. S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter), 1079–1098.
- Busso, C., Deng, Z., Neumann, U., and Narayanan, S. (2005). Natural head motion synthesis driven by acoustic prosodic features. *J. Visual. Comput. Animat.* 16, 283–290. doi: 10.1002/cav.80
- Calbris, G. (2011). *Elements of Meaning in gesture*. Amsterdam: John Benjamins Publishing.
- Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., et al. (2006). "The AMI meeting corpus: a pre-announcement," in *Proceedings of the Second International Conference on Machine Learning for Multimodal Interaction, MLMI'05* (Berlin; Heidelberg: Springer-Verlag), 28–39.
- Cassell, J. (2000). *Embodied Conversational Agents*. Cambridge, MA: MIT Press.
- Chiu, C.-C., and Marsella, S. (2014). "Gesture generation with low-dimensional embeddings," in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems* (Paris: International Foundation for Autonomous Agents and Multiagent Systems), 781–788.
- Chiu, C.-C., Morency, L.-P., and Marsella, S. (2015). "Predicting co-verbal gestures: a deep and temporal modeling approach," in *International Conference on Intelligent Virtual Agents* (Delft: Springer), 152–166.
- Chui, K. (2011). Conceptual metaphors in gesture. *Cogn. Linguist.* 22, 437–458. doi: 10.1515/cogl.2011.017
- Cienki, A. (1998). "Metaphoric gestures and some of their relations to verbal metaphoric expressions," in *Discourse and Cognition: Bridging the Gap*, 189–204.
- Cienki, A. (2005). "Image schemas and gesture," in *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, Vol. 29, ed B. Hampe (Berlin; New York, NY: Mouton de Gruyter), 421–442.
- Cienki, A. (2013). "Image schemas and mimetic schemas in cognitive linguistics and gesture studies," in *Review of Cognitive Linguistics. Published Under the Auspices of the Spanish Cognitive Linguistics Association*, Vol. 11, ed J. Benjamins (Amsterdam: John Benjamins), 417–432.
- Cienki, A., and Müller, C. (2008). *Metaphor and Gesture*. Amsterdam: John Benjamins Publishing.
- Clausner, T. C., and Croft, W. (1999). Domains and image schemas. *Cogn. Linguist.* 10, 1–32.
- Croft, W., and Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge, UK: Cambridge University Press.
- Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *J. Pers. Soc. Psychol.* 23:283. doi: 10.1037/h0033031
- Durupinar, F., Kapadia, M., Deutsch, S., Neff, M., and Badler, N. I. (2016). Perform: perceptual approach for adding ocean personality to human motion using laban movement analysis. *ACM Trans. Graph.* 36:6. doi: 10.1145/3072959.2983620
- Eyben, F., Weninger, F., Gross, F., and Schuller, B. (2013). "Recent developments in opensmile, the munich open-source multimedia feature extractor," in *Proceedings of the 21st ACM International Conference on Multimedia, MM '13* (New York, NY: ACM), 835–838.
- Grady, J. E. (2005). "Image schemas and perception: refining a definition," in *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, ed B. Hampe (Berlin; New York, NY: Walter de Gruyter GmbH & Co. KG), 35–55.
- Gris, I., Rivera, D. A., Rayon, A., Camacho, A., and Novick, D. (2016). "Young merlin: an embodied conversational agent in virtual reality," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction, ICMI 2016* (New York, NY: ACM), 425–426.
- Halliday, M., Matthiessen, C. M., and Matthiessen, C. (2014). *An Introduction to Functional Grammar*. Abingdon: Routledge.
- Harvey, P. H., Currie, E., Daryanani, P., and Augusto, J. C. (2015). "Enhancing student support with a virtual assistant," in *International Conference on E-Learning, E-Education, and Online Training* (Novedrate: Springer), 101–109.
- Heylen, D., Kopp, S., Marsella, S. C., Pelachaud, C., and Vilhjálmsdóttir, H. (2008). "The next step towards a function markup language," in *Intelligent Virtual Agents*, eds B. Willem-Paul, J. Broekens, and D. Heylen (Delft: Springer), 270–280.

- Hostetter, A. B., and Alibali, M. W. (2008). Visible embodiment: gestures as simulated action. *Psychon. Bull. Rev.* 15, 495–514. doi: 10.3758/PBR.15.3.495
- Hostetter, A. B., and Alibali, M. W. (2010). Language, gesture, action! A test of the Gesture as simulated action framework. *J. Mem. Lang.* 63, 245–257. doi: 10.1016/j.jml.2010.04.003
- Johnson-Laird, P. N. (2006). *How We Reason*. Oxford University Press.
- Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago, IL: University of Chicago Press.
- Kendon, A. (1980). "Gesture and speech: two aspects of the process of utterance," in *Nonverbal Communication and Language*, ed M. R. Key (The Hague: Mouton), 207–227.
- Kendon, A. (2000). "Language and gesture: unity or duality?" in *Language and Gesture, Language Culture and Cognition*, Vol. 2. ed D McNeill (Cambridge University Press), 47–63.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge, UK: Cambridge University Press.
- Kipp, M., Neff, M., Kipp, K. H., and Albrecht, I. (2007). "Towards natural gesture synthesis: evaluating gesture units in a data-driven approach to gesture synthesis," in *International Workshop on Intelligent Virtual Agents* (Paris: Springer), 15–28.
- Kopp, St., Tepper, P., Ferriman, K., Striegnitz, K., and Cassell, J. (2007). "Trading Spaces: how humans and humanoid use speech and gesture to give directions," in *Conversational Informatics: An Engineering Approach*, ed T. Nishida (John Wiley & Sons), 133–160.
- Lakoff, G., and Johnson, M. (1980). Conceptual metaphor in everyday language. *J. Philos.* 77, 453–486. doi: 10.2307/2025464
- Lee, J., and Marsella, S. (2006). "Nonverbal behavior generator for embodied conversational agents," in *International Workshop on Intelligent Virtual Agents* (Berlin/Heidelberg: Springer), 243–255.
- Lee, J., and Marsella, S. (2009). "Learning a model of speaker head nods using gesture corpora," in *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, Vol. 1 (Amsterdam: International Foundation for Autonomous Agents and Multiagent Systems), 289–296.
- Leonard, T., and Cummins, F. (2011). The temporal relation between beat gestures and speech. *Lang. Cogn. Process.* 26, 1457–1471. doi: 10.1080/01690965.2010.500218
- Lesk, M. (1986). "Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone," in *Proceedings of the 5th Annual International Conference on Systems Documentation* (Toronto, ON: ACM), 24–26.
- Levine, S., Krähenbühl, P., Thrun, S., and Koltun, V. (2010). "Gesture controllers," in *ACM Transactions on Graphics (TOG)*, Vol. 29. (New York, NY: ACM), 124.
- Levine, S., Theobalt, C., and Koltun, V. (2009). "Real-time prosody-driven synthesis of body language," in *ACM Transactions on Graphics (TOG)*, Vol. 28. (New York, NY: ACM), 172.
- Lhommet, M., and Marsella, S. (2014). "Metaphoric gestures: towards grounded mental spaces," in *International Conference on Intelligent Virtual Agents* (New York, NY: Springer), 264–274.
- Lhommet, M., and Marsella, S. (2016). "From embodied metaphors to metaphoric gestures," in *Proceedings of Annual Meeting of the Cognitive Science Society* (Philadelphia, PA).
- Lhommet, M., Xu, Y., and Marsella, S. (2015). "Cerebella: automatic generation of nonverbal behavior for virtual humans," in *AAAI* (Austin, TX), 4303–4304.
- Lisetti, C., Amini, R., and Yasavur, U. (2015). Now all together: overview of virtual health assistants emulating face-to-face health interview experience. *KI-Künstliche Intelligenz* 29, 161–172. doi: 10.1007/s13218-015-0357-0
- Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Lab. Phonol.* 3, 71–89. doi: 10.1515/lp-2012-0006
- Lücking, A., Mehler, A., Walther, D., Mauri, M., and Kurfürst, D. (2016). "Finding recurrent features of image schema gestures: the figure corpus," in *Proceedings of the 10th International Conference on Language Resources and Evaluation. LREC* (Portorož).
- Marsella, S., Xu, Y., Lhommet, M., Feng, A., Scherer, S., and Shapiro, A. (2013). "Virtual character performance from speech," in *Symposium on Computer Animation* (Anaheim, CA).
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychol. Rev.* 92, 350–371. doi: 10.1037/0033-295X.92.3.350
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Cambridge, UK: University of Chicago Press.
- McNeill, D., and Levy, E. (1980). *Conceptual Representations in Language Activity and Gesture*. Hoboken, NJ: ERIC Clearinghouse.
- Mehler, A., Lücking, A., and Abrami, G. (2015). Wikinet: Image schemata as a basis of gestural writing for kinetic museum wikis. *Univ. Access Inform. Soc.* 14, 333–349. doi: 10.1007/s10209-014-0386-8
- Miller, G. A. (1995). Wordnet: A lexical database for english. *Commun. ACM* 38, 39–41. doi: 10.1145/219717.219748
- Mittelberg, I. (2008). Peircean semiotics meets conceptual metaphor: iconic modes in gestural representations of grammar. *Metaphor Gesture* 3, 115–154. doi: 10.1075/g3.3.08mit
- Morton, T., Kottmann, J., Baldridge, J., and Bierner, G. (2005). *Opennlp: A Java-based nlp Toolkit*. Available online at: <http://opennlp.sourceforge.net>
- Nadeau, D., and Sekine, S. (2007). A survey of named entity recognition and classification. *Lingvist. Investigat.* 30, 3–26. doi: 10.1075/li.30.1.03nad
- Pecune, F., Cafaro, A., Chollet, M., Philippe, P., and Pelachaud, C. (2014). "Suggestions for extending saiba with the vib platform," in *Workshop Architectures and Standards for IVAs, International Conference of Intelligent Virtual Agents* (Los Angeles, CA), 16–20.
- Poggi, I. (2007). *Mind, Hands, Face and Body: A Goal and Belief View of Multimodal Communication*. Körper, Zeichen, Kultur. (Berlin: Weidler).
- Reddy, M. (1979). The conduit metaphor. *Metaphor Thought* 2, 164–201.
- Sargin, M. E., Yemez, Y., Erzincan, E., and Tekalp, A. M. (2008). Analysis of head gesture and prosody patterns for prosody-driven head-gesture animation. *IEEE Trans. Pattern Anal. Mach. Intell.* 30, 1330–1345. doi: 10.1109/TPAMI.2007.70797
- Terken, J. (1991). Fundamental frequency and perceived prominence of accented syllables. *JASA* 89, 1768–1776. doi: 10.1121/1.401019
- Thiebaux, M., Marsella, S., Marshall, A. N., and Kallmann, M. (2008). "Smartbody: behavior realization for embodied conversational agents," in *Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems*, Vol. 1. (Estoril: International Foundation for Autonomous Agents and Multiagent Systems), 151–158.
- Toutanova, K., Klein, D., Manning, C. D., and Singer, Y. (2003). "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, Vol. 1. (Edmonton, AB: Association for Computational Linguistics), 173–180.
- Tversky, B., and Hard, B. M. (2009). Embodied and disembodied cognition: spatial perspective-taking. *Cognition* 110, 124–129. doi: 10.1016/j.cognition.2008.10.008
- Vilhjálmsdóttir, H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S., et al. (2007). "The behavior markup language: recent developments and challenges," in *International Workshop on Intelligent Virtual Agents* (Paris: Springer), 99–111.
- Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction: an overview. *Speech Commun.* 57, 209–232. doi: 10.1016/j.specom.2013.09.008
- Wang, Y., and Neff, M. (2013). "The influence of prosody on the requirements for gesture-text alignment," in *International Workshop on Intelligent Virtual Agents* (Edinburgh, UK: Springer), 180–188.
- Wilson, A., and Golonka, S. (2013). Embodied cognition is not what you think it is. *Front. Psychol.* 4:58. doi: 10.3389/fpsyg.2013.00058
- Xu, Y., Pelachaud, C., and Marsella, S. (2014). "Compound gesture generation: a model based on ideational units," in *International Conference on Intelligent Virtual Agents* (Boston, MA: Springer), 477–491.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ravenet, Pelachaud, Clavel and Marsella. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Prosodic Characteristics of Non-referential Co-speech Gestures in a Sample of Academic-Lecture-Style Speech

Stefanie Shattuck-Hufnagel* and Ada Ren

Speech Communication Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, United States

OPEN ACCESS

Edited by:

Wendy Sandler,
University of Haifa, Israel

Reviewed by:

Daniel Loehr,
Georgetown University, United States
Pilar Prieto,
Institució Catalana de Recerca i
Estudis Avançats (ICREA), Spain

*Correspondence:

Stefanie Shattuck-Hufnagel
sshuf@mit.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 16 January 2018

Accepted: 31 July 2018

Published: 07 September 2018

Citation:

Shattuck-Hufnagel S and Ren A
(2018) The Prosodic Characteristics
of Non-referential Co-speech
Gestures in a Sample of
Academic-Lecture-Style Speech.
Front. Psychol. 9:1514.
doi: 10.3389/fpsyg.2018.01514

Many studies have documented a close timing relationship between speech prosody and co-speech gesture, but some studies have not, and it is unclear whether these differences in speech-gesture alignment are due to different speaking tasks, different target gesture types, different prosodic elements, different definitions of alignment, or even different languages/speakers. This study contributes to the ongoing effort to elucidate the precise nature of the gesture–speech timing relationship by examining an understudied variety of American English, i.e., academic-lecture-style speech, with a focus on an understudied type of gesture: Non-Referential gestures, which make up the majority of this corpus. Results for the 1,334 Stroke-Defined Gestures in this 20-min sample suggest that the stroke phase of a Non-Referential gesture tends to align with a pitch-accented syllable, just as reported in studies of other gesture types (e.g., deictic gestures) and in other speaking styles (such as narration). Preliminary results are presented suggesting that trajectory shapes of these Non-Referential gestures are consistent across a higher-level prosodic grouping, supporting earlier proposals for kinematic constancy across spoken prosodic constituents (Kendon, 1972, 1980, 2004). Analysis also raises the possibility that the category of Non-Referential gestures is not solely made up of ‘beats,’ defined as simple bi-phasic flick-like movements that beat out the rhythm of the speech, but includes gestures with multiple phases and various types of rhythmicity. Taken together, the results of this analysis suggest (1) a wide range of gesture configurations within the undifferentiated category of Non-Referential gestures or ‘beats,’ which requires further investigation, and (2) a close coordination between co-speech gestures and the prosodic structure of spoken utterances across speaking styles and gesture referentiality, which has profound implications for modeling the process of planning an utterance.

Keywords: co-speech gesture, speech prosody, speech production planning, prosodic prominence, prosodic constituents

INTRODUCTION

The relationship between spoken utterances and the co-speech gestures that often accompany them has been the subject of great interest over the centuries, and this interest has intensified with the development of modern prosodic theory. Over the past few decades, the incorporation of phrase-level prosodic constituency and prominence patterns into linguistic grammars (e.g., Liberman and Prince, 1977; Selkirk, 1984; Beckman and Pierrehumbert, 1986; Nespor and Vogel, 1986), along with the development of an extensive system for capturing significant systematic aspects of gestural movements and their communicative function (Kendon, 1972, 1980, 2004; McNeill, 1992, 2005) has opened the door to a range of studies asking how these two streams of behavior interact. This is an important question, because to the extent that both sets of actions contribute to the communication of a message during the act of speaking, it is a reasonable presumption that they are planned together (Esteve-Gibert and Prieto, 2013; Krivokapic, 2014; Wagner et al., 2014; Krivokapic et al., 2015; Esteve-Gibert et al., 2017). Such a view has critical implications for the development of a comprehensive model of the speech production planning process. Moreover, evidence that the two sets of actions are closely timed with respect to each other has the potential to implicate a prosodic representation as the integrating planning framework for both speech articulation and co-speech gesture, since it is increasingly apparent that prosodic structure is one of the major factors governing speech timing (Wightman et al., 1992; Byrd et al., 2006; Turk and Shattuck-Hufnagel, 2007 *inter alia*).

In his influential 1992 book *Hand and Mind*, David McNeill proposed a categorization scheme for co-speech gestures that separated Referential gestures, which visually illustrate some aspect of the semantic content of the speech they accompany, from Non-Referential gestures, which can be said to convey information about the form of the utterance rather than its content. Referential gestures have been subdivided into various types (McNeill, 1992), including iconic (illustrating concrete aspects of the speech content), metaphoric (illustrating abstract aspects), and deictic (pointing to actual or symbolic locations). These Referential gesture subclasses, based on the philosophical work of Pierce (1960; McNeill and Levy, 1982), have understandably been of particular interest, because their relationship to the meaning of the speech is often straightforward to identify. Moreover, this relationship is compelling as an argument for the integration of the planning for speech and gestural movements as co-signaling systems for the meaning of an intended message. Thus, Referential gestures have been extensively explored in a wide range of studies, which have revealed their striking contribution to acts of communication. In contrast, Non-Referential gestures, which are often called ‘beats’ (or sometimes, ‘batons,’ Efron, 1941/1972), have not been as extensively subcategorized. The term ‘beats’ suggests a degree of rhythmic periodicity, invoking a conductor beating out the rhythm of an orchestral performance, and Non-Referential gestures have sometimes been defined in these terms, as e.g., beating out the rhythm of the speech. Alternatively, McNeill (1992) describes a particular kind of beat, i.e., a single

in-out or up-down flick of the finger or hand, which he notes can mark particular locations in a narrative structure. But Non-Referential gestures or beats have been primarily defined as ‘not iconic, metaphoric or deictic,’ leaving a substantial gap in our understanding of the range of behaviors in this set of gestural movements.

Although timing with respect to spoken prosody has been particularly important for Non-Referential gestures or beats, because they have been defined in terms of their relationship to the rhythm of speech, i.e., to the pattern of prominences in an utterance, Referential gestures have also been described as temporally aligned with the prosodic structure of speech. For example, Kendon (1972, 1980) proposed a hierarchy of prosodic units, from tone groups to locutions, locution groups, locution clusters and the discourse, and a corresponding hierarchy of gestural structures, from gesture phrases to gestural units. In a short sample of videoed conversation that he analyzed in great detail, he reported that these two sets of units were closely coordinated, such that, e.g., gesticular movements in successive tone groups differ in some characteristics, while sharing other characteristics if they formed a larger constituent, a locution (generally a full sentence). He noted that co-speech gestures may illustrate objects or actions referred to in the speech, or they may indicate the organizational structure of the elements of the discourse. Thus he did not distinguish sharply between Referential gestures that visually illustrate an aspect of the speech, and Non-Referential gestures that have other functions, in their likelihood of aligning with prosodic structure.

Other investigators who have focused on gesture-prosody alignment have also looked at co-speech gestures as a single category, without distinguishing between Referential and Non-Referential categories. For example, Loehr (2004) reports temporal alignment between gestural strokes and spoken pitch accents (i.e., phrase-level prominences signaled by F0), without distinguishing among gesture types, and Shattuck-Hufnagel et al. (2007) report similar findings for gestures with sudden sharp end points (which they called ‘hits’). Investigations have sometimes focussed on the alignment of particular subtypes of Referential gestures, particularly deictic or pointing movements, and eyebrow movements (Krahmer et al., 2002; Keating et al., 2003), that appear to have a prominence-lending function. Thus the question of how Non-Referential gestures, as a specific subset of co-speech gestures, align with spoken prosody has not been thoroughly investigated. This paper reports some preliminary results from a larger study aimed at extending our current understanding of the relationship between the prosody of a spoken utterance and Non-Referential co-speech gestures in an understudied speaking style, i.e., formal academic lectures. Initial informal observation suggested that this style elicits a large proportion of Non-Referential gestures, which also provides an opportunity to begin to examine the range and structure of this category of gestures, which appears not to be homogeneous. Thus the research questions addressed in this paper are (1) Are Non-Referential gestures the predominant type in this speech sample? (2) Do the Non-Referential gestures exhibit alignment with prosodic structure? And (3) Are Non-Referential gestures a

homogeneous category, as suggested by their designation as 'beats'?

MATERIALS AND METHODS

In the course of designing and carrying out this study, two issues have come to the fore. The first concerns the question of how to convey, in visual terms, the path of a gestural movement. Many different conventions have been used in the literature to capture on the printed page the dynamic aspects of a movement, which the viewer can easily discern when watching the speaker in person or watching a video. But none of these existing conventions seemed precisely satisfactory for our purposes. To supplement these conventions, we have developed a tool called the gestural sketch, which is a simple line drawing of the path that the hand traverses during a continuous sequence of gestures. This tool plays an important role in describing the degree of similarity or dissimilarity between successive gestures, as well as the trajectory shapes for gesture sequences that are perceived as beatlike.

The second issue concerns the size of the prosodic constituent in the speech that is most useful for reporting our results on gesture grouping. The prosodic hierarchy for spoken utterances is generally taken to have the Utterance as its highest constituent. Each Utterance is made up of one or more Full Intonational Phrases (marked by a Boundary Tone on the final syllable), with each Full Intonational Phrase made up of one or more Intermediate Intonational Phrases (marked by at least one Pitch Accent (phrase-level prominence) and a Phrase Tone controlling the fundamental frequency contour between the final pitch accent and the end of the phrase), and so on down the hierarchy (see Shattuck-Hufnagel and Turk, 1996 for a summary). On this view there are clear definitional characteristics that permit the identification of Full and Intermediate Intonational Phrases in the signal, but it is less clear what marks the edges of an Utterance or of even higher-level constituents in the hierarchy, such as the Locution or the Discourse (see Kendon, 1980 for discussion). This problem was addressed here by extending an existing method for prosodic annotation called Rapid Prosodic Transcription (RPT), which was developed by Cole et al. (2010). Extending the RPT method produced a 'crowd-sourced' identification of the higher-level constituents that were required for our study.

Although studies of the temporal alignment of co-speech gestures have generally found a close relationship between gesture timing and prosodic timing, this is not always the case (McClave, 1994; Ferre, 2005, 2010). This raises the question of whether different types of gestures and/or different types of speaking show different patterns of alignment. However, it is difficult to address this question because different studies have looked at different speaking tasks (e.g., spontaneous conversational speech, emotional speech, speech elicited in the laboratory via highly constrained tasks) and different types of gestures (often deictic), as well as different parts of a gesture [e.g., the gesture stroke defined as the high-intensity movement to a target (Yasinnik et al., 2004)] vs. the gesture stroke as the period during which the arm maintains its maximum extension in a

deictic gesture, vs. the apex (Jannedy and Mendoza-Denton, 2005) and different locations in the spoken prosody (e.g., the discrete point of maximum F0 for a high pitch accent, vs. the time interval of the accented syllable). In addition, different speakers appear to produce different proportions of gesture types (Myers, 2012, described in Krivokapic, 2014), and findings appear to differ across languages. This rich variety in sample materials and methodological approaches and results has resulted in a range of findings (see Krivokapic, 2014; Wagner et al., 2014 for reviews) that suggest the need for a comprehensive comparison of timing patterns across speaking tasks, gesture types and prosodic structures, to determine the generalizability of individual findings. Such comprehensive coverage is a very long-term project; the study described in this paper contributes to this long-term goal by focussing on a sample of an understudied speaking style (academic lectures) in which the gestures are predominantly Non-Referential. The analysis includes alignment of the gestural strokes both with prosodic prominences and, in a preliminary way, with higher-level prosodic constituents. Results point the way to further studies to elucidate how speech and co-speech gestures interact in a communicative event, and they suggest some constraints on the set of appropriate models of the planning process that produces such an event.

The analyses carried out in this study required the hand-labeling of a wide range of characteristics of both the speech and the co-speech gestures. This labeling process provides the information that is necessary in order to test hypotheses about how these two streams of behavior are aligned with each other, and will be described in some detail.

The Corpus

The availability of a video-recorded speech sample that provides a high proportion of Non-Referential gestures was discovered by accident, when a set of commercially available academic lectures (available from The Teaching Company/Great Courses Company¹) was chosen as an object of study. According to information provided by the company, these lecturers are selected for their popularity on their respective campuses, and recruited to deliver a course in half-hour lectures to a small audience that is physically present in the room. The lectures are recorded on video and offered for sale to the public. It can be presumed that the lecturers selected in this way are effective communicators, and in our experience they generally produce fluent speech as well as large numbers of co-speech gestures. The sound quality of the recordings is also high, facilitating transcription of the utterances as well as annotation of their prosody. These videos were originally selected for the study of gesture-prominence alignment in part because they provide a clear view of the speaker's upper body (**Figure 1**), which is filmed directly from the front. As a result, most of the time it is possible to view the full extent of the hands, arms, head and upper torso (at least when the speaker is not occluded by an illustrative graphic). In addition, these highly practiced college professors produce their lectures quite fluently, so that prosodic analysis of the prominences and word groupings in their

¹<https://www.thegreatcourses.com/>

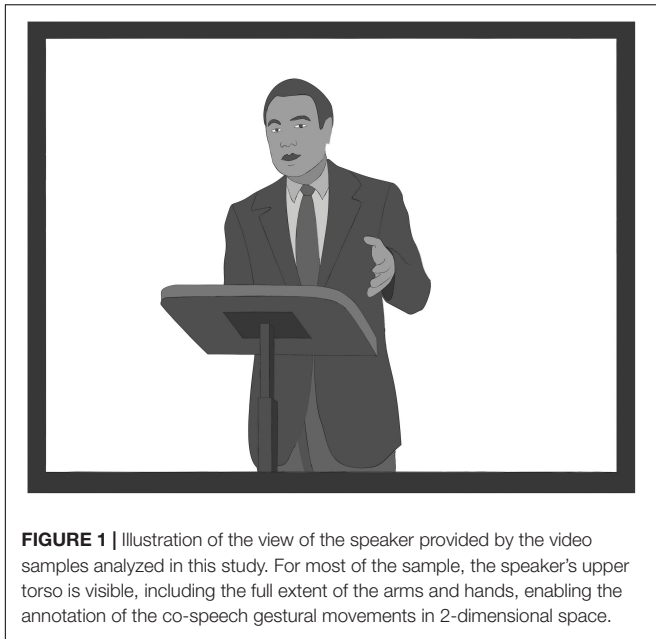


FIGURE 1 | Illustration of the view of the speaker provided by the video samples analyzed in this study. For most of the sample, the speaker's upper torso is visible, including the full extent of the arms and hands, enabling the annotation of the co-speech gestural movements in 2-dimensional space.

speech is less challenging than for more typical speech, with its hesitations, restarts and other disfluencies. As we began to label the temporal locations of the gestures in these videos, we noticed that a large proportion of the gestures did not appear to be Referential, in the sense of visually illustrating the content of the accompanying speech in any obvious way. Thus was born the idea of analyzing this set of gestures with respect to its alignment with spoken prosody, in order to compare the results with existing observations of these alignment patterns for gestures which were either explicitly Referential or not distinguished with respect to their referential nature.

The subsample from the larger study that will be discussed in this paper includes an entire 30-min lecture produced by one male speaker (here referred to as the London sample, after its topic), which comprises 30 min 47 s of speech, with 23 min 35 s of video useable for gesture analysis. (For the excluded 7 min 12 s the speaker was not visible due to the display of illustrative graphics. The word transcriptions and prosody of the excluded portions of the sample are available for future analysis of the discourse structure of the lecture.)

Labeling

In this paper, we focus on the manual co-speech gestures in the sample, i.e., those that involve the hand(s) and arm(s) of the speaker [The co-speech movements of other articulators, such as the head, eyebrows, direction of gaze and upper torso, are also of interest for their alignment with spoken prosody (e.g., McClave, 2000; Keating et al., 2003; Shattuck-Hufnagel et al., 2010; Swerts and Krahmer, 2010; Esteve-Gibert et al., 2017) but will not be discussed here]. For most aspects of the labeling, the gestures were annotated without listening to the speech, and the speech without viewing the video, to avoid any possibility of the labeler's judgment about events in one channel being influenced by events in the other. However, this was not possible for one type

of annotation, i.e., determining the referentiality of the co-speech gestures; this required the labeler to listen and look at the same time, because the decision depends on the relationship of the gestures to the meaning of the speech. Unless otherwise noted, each type of gestural annotation described below was carried out while viewing the silent video, and each type of speech annotation while listening to the sound recording only.

Gesture Annotation

The core of this study concerns the annotation of meaningful co-speech gestures, based on an exhaustive analysis of all movements made by the speaker during a lecture. Such annotation is not a trivial matter. Movements which occur during the speech, but can be regarded as not planned to be part of a communicative act, must be identified as such, and distinguished from intentional movements that appear to be part of a communicative act. These include movements such as grasping the podium, reaching out to turn a page, or very small 'drifting' movements of the hands or fingers, in which the articulator moves slowly in space in what appears to be a non-directed way. [There are gestures that can be interpreted as information by listeners, such as self-grooming actions like tucking the hair behind the ear or tossing the head, and may even in some cases be planned by the speaker to communicate information (such as flirtatiousness), but such movements are not included here. The question of whether even movements in this category are aligned with the spoken prosody is left for another day.] For the purposes of this study, we define movements planned to be part of a communicative act as those which include a stroke, i.e., an intentional movement that is sometimes referred to as 'the business portion' of a co-speech gesture. Thus the first annotation step was to identify the set of movements that each include a stroke, i.e., the set of Stroke-Defined Gestures (SDGs); this category defines the set of movements analyzed in this study. All of the gesture annotations were carried out by the second author, who is highly experienced in gesture labeling. Additional information on the suite of labeling methods can be found at <http://adainspired.mit.edu/gesture-research/>.

Gesture strokes

Gesture strokes were identified using the ELAN annotation software². As noted above, strokes were distinguished from a number of other movement types, such as small undirected movements that lacked a sense of intentionality, task-related movements and drifting movements. The time of onset and offset of each stroke movement was annotated in a Stroke tier in ELAN.

Once the strokes are identified, specifying the Stroke-Defined Gestures (SDGs), a number of additional labeling steps can be carried out. Annotation results that will be reported here include the Referentiality of the gesture, its handedness, and for sequences of gestures, their perceived grouping. A number of additional characteristics have also been labeled for this sample, including, e.g., the optional gesture phases (i.e., preparation, pre-stroke hold, post-stroke hold and recovery, as proposed in Kendon, 1980, 2004); handshape (and change in handshape); trajectory shape (straight, curved or looping, i.e., forming a closed

²<https://tla.mpi.nl/tools/tla-tools/elan/>

curve); and location with respect to the speaker's body; these results will be reported in a subsequent publication and will not be discussed further here.

Gesture referentiality

Referentiality was labeled for each Stroke-Defined-Gesture, using an annotation scheme that included two categories: Referential and Non-Referential. As noted above, this labeling task (unlike the remaining tasks) was carried out while both viewing the video and listening to the speech.

Gesture handedness

Gesture handedness was labeled with an annotation scheme that included Left and Right for gestures made with one hand; two-handed gestures were labeled as Bimanual-synchronous or Bimanual-asynchronous (i.e., the two hands do not produce symmetrical movements), and Bimanual-L-dominant or Bimanual-R-dominant.

Perceived gesture groupings

As part of the larger ongoing study, sequences of gestures that were perceived as occurring in a group were labeled as Perceived Gesture Groups (PGGs), while looking at the silent video. This terminology was adopted instead of Kendon's 'Gesture Units,' because Gesture Units are proposed to conclude with a relaxation phase, and it was not yet certain that the gesture sequences perceived as grouped had this characteristic. These PGGs formed the basis for analysis of the corresponding Gesture Sketches described below.

Gesture sketches

Gesture sketches were developed to provide a visual impression of the trajectory of a gestural movement, and to explore the possibility that they could facilitate judging the similarity in this characteristic across a sequence of gestures. Gesture sketches are line drawings of the trajectory through space of the moving hand, illustrated in **Figure 2** above. They provide a more detailed sense of the sometimes-complex path of movement of the hand through space than is possible using either a single-word characterization of path shape, or a short arrow added to a drawing of the speaker indicating direction of movement. They do not capture additional aspects of the movement, such as its timing, changes in velocity over time or its alignment with the speech, but for the purposes of this study they have proven to be a useful indicator.

Additional gesture characteristics and components that have been annotated, but are not discussed in this paper include gesture phases (Kendon, 1980, 2004), handshape, trajectory shape, and location with respect to the speaker's body; these labels are designed to facilitate the quantitative estimation of similarity between one gesture and the next, and to test hypotheses about the cues to gesture grouping (Shattuck-Hufnagel and Ren, 2012). For some of the samples in the larger study, movements of the head, eyebrows and upper torso have also been annotated (Shattuck-Hufnagel et al., 2010), to facilitate comparison of the prosodic timing of gestures by various articulators, and investigation of the coordination among co-speech movements of various body parts.

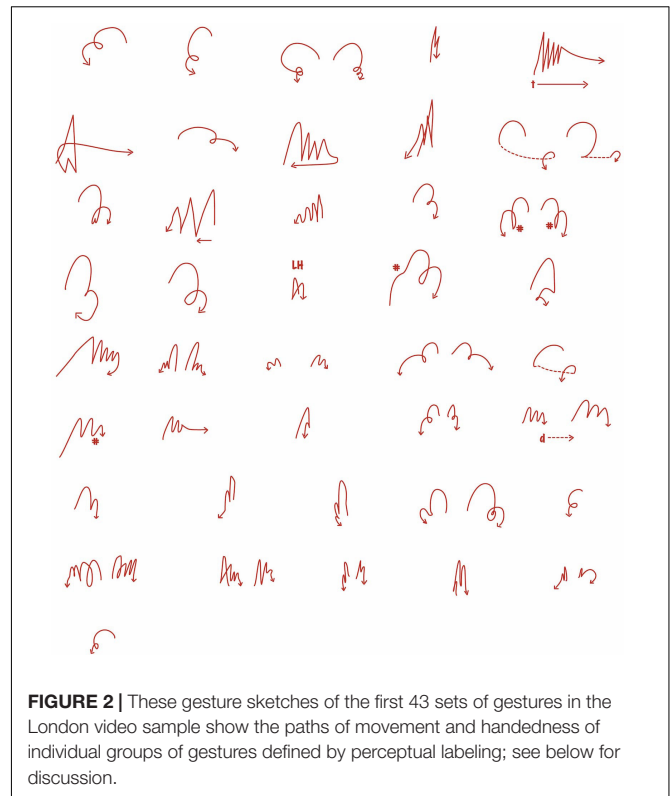


FIGURE 2 | These gesture sketches of the first 43 sets of gestures in the London video sample show the paths of movement and handedness of individual groups of gestures defined by perceptual labeling; see below for discussion.

Speech Annotation

To determine the timing relationship between the co-speech gestures and the prosodic constituent and prominence structure of the speech, the speech was transcribed orthographically and labeled for its intonational structure using Praat³ as a display and labeling tool and ToBI⁴ as the prosodic annotation system. This annotation was carried out by the first author, who is an experienced ToBI labeler. ToBI labels include, among other prosodic characteristics, the nature and location of tonal targets that signal phrase-level prosodic prominences (pitch accents) and two levels of intonational phrasing: higher-level Full Intonational Phrases, and the lower-level Intermediate Intonational Phrases that make up the higher level phrases.

In addition, to facilitate analysis of the temporal overlap between gestural strokes and pitch-accented syllables, a rough segmentation of the speech wave form into syllables was carried out. This segmentation task is challenging for utterances in English, where the syllable affiliation of an inter-vocalic consonant in words like *movie* or *label* is not always clear, but an approximate segmentation was carried out despite this difficulty.

Finally, when initial analyses made it clear that a larger prosodic constituent than the Full Intonational Phrase would be necessary in order to reach a clearer understanding of the relationship between the grouping of successive gestures and the prosodic constituent structure of the speech, a method for

³<http://www.fon.hum.uva.nl/praat/>

⁴<https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/>

annotating the higher-level prosodic constituents was developed. This approach uses an extended version of the RPT, developed by Cole et al. (2010, 2014), in which untrained participants listen to recorded utterances and mark them (in real time) for prominence and word-group boundaries. In RPT, listeners mark only one size or level of boundary, and the annotations of multiple listeners are summed to provide a continuous-valued estimate of the size or level of the constituent boundary. While it is unclear exactly what criteria listeners use in determining the location of the boundaries they mark, since the full range of semantic and syntactic as well as prosodic information is available to them, Cole et al. (2010, 2014) have reported good agreement among listeners and a correlation of Rapid-Prosodic-Transcription-defined groupings with intonational phrases and prominences annotated by highly trained ToBI labelers. We extended this method by inviting listeners to mark three levels of constituent boundary rather than just one level, using a single slash (/) for the smallest grouping, a double slash (//) for a deeper boundary of a higher-level grouping, and a triple slash (///) for the deepest boundary of the highest-level grouping. As in RPT, Extended RPT boundary markers are then summed across listeners, to provide an estimate of the perceived higher-level word groupings.

RESULTS

The analyses reported in this paper address three specific questions about the Stroke-Defined Gestures in the sample. First, is there a high proportion of Non-Referential gestures in this sample, as our preliminary impression suggested. Second, how do the strokes of these Non-Referential gestures align with the prosodic prominences of the speech they accompany. And third, do these Non-Referential gestures form a unified class. Before turning to these questions, we first summarize some of the characteristics of this sample.

Corpus Characteristics

The 23-min portion of video in which the speaker was not occluded by graphics was labeled with 1,334 Stroke-Defined Gestures. The speech that accompanied these non-occluded regions was labeled with 2,065 Pitch Accent labels, 682 Full Intonational Phrase labels (ToBI Break Index 4), and 978 Intermediate Intonational Phrase labels (ToBI Break Index 3).

Are Most of the Stroke-Defined-Gestures in This Corpus Non-referential?

Of the 1,334 SDGs identified in this sample, 1,263 (94.6%) were labeled as unambiguously Non-Referential, and 70 (5.4%) as

Referential. (One gesture overlapped with a non-speech region and was omitted from further analysis.) This result confirms our initial informal impression that most of the manual co-speech gestures employed by this speaker are not referential. To our knowledge, extensive tabulations of the proportion of Referential vs. Non-Referential gestures are not available in the literature, so it is not yet possible to determine whether this proportion is atypically large. However, it appears that for this speaker, speaking in this style or circumstance, Non-Referential gestures predominate. This provides an opportunity to determine whether these largely Non-Referential co-speech gestures align with the prominent syllables of the speech, just as has been reported for individual Referential gestures and for corpora of gestures not sorted by their referentiality.

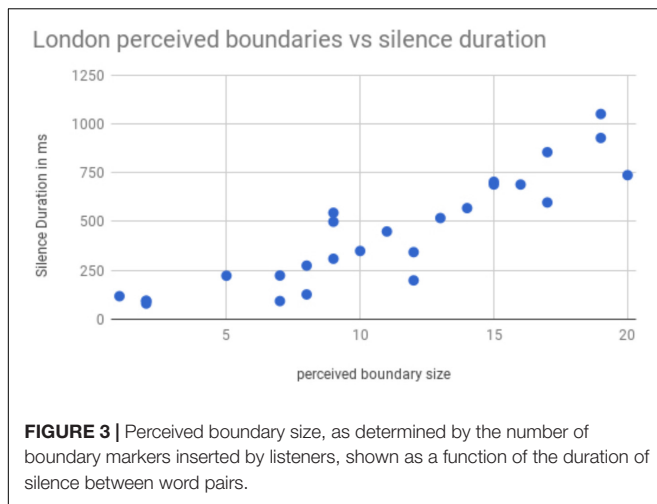
Do These Gestural Strokes Align With Spoken Prominences?

In this study, alignment between a Stroke-Defined Gesture and a spoken prominence was defined as any degree of overlap between the temporal region labeled as an accented syllable and the region labeled as the gestural stroke. Although this definition of an association between strokes and accented syllables is more stringent than some in the literature, results are nevertheless consistent with earlier reports (Loehr, 2004, 2012; Renwick et al., 2004; Shattuck-Hufnagel et al., 2007), in that the proportion of strokes in gestures perceived as Non-Referential in this speech sample that overlap in time with accented syllables is very high: 83.1% (Table 1). This proportion does not differ substantially from that for the (much smaller number of) gestures perceived as Referential.

This result suggests that, like the strokes of Referential gestures, the strokes of Non-Referential gestures tend to occur in conjunction with spoken prominences. That is perhaps unsurprising, in view of the general understanding of Non-Referential gestures as ‘beats’ which mark out the rhythm of the speech they accompany—but recall that these gestures were annotated from the video alone, without access to the accompanying speech. Thus this high percentage of overlap raises several interesting questions about how two types of alignment between spoken prominence and gestural stroke are related. On the one hand, the stroke of a Referential gesture is often aligned with a phrasally prominent syllable (see Kendon, 2004; Ch. 7), and on the other hand, the strokes of Non-Referential gestures in this sample are also reliably aligned with pitch-accented syllables. In discussions of the alignment of Referential-gesture strokes with phrasally prominent syllables, little mention is made of concepts such as ‘beating out the rhythm of the accompanying speech,’ whereas in discussion of the alignment of

TABLE 1 | The proportion of SDGs whose strokes overlap in time with a pitch-accented syllable, for gestures perceived as Referential vs. Non-Referential.

	Tokens with overlap of stroke w/PAcc syllable	Tokens with no overlap of stroke w/PAcc syllable	Total	Percent of tokens that overlap with a PAcc syllable
Referential	58	12	70	82.85%
Non-Referential	1,050	213	1,263	83.13%
Total	1,108	225	1,333	83.12%



Non-Referential gesture strokes (or beats), this characterization is common. Future work will need to sort out whether the alignment phenomena for these two sets of co-speech gestures have a mechanism in common, or whether the alignment of Non-Referential gestures with the prominence patterns of the speech is, for example, more reliable in regions where the spoken prominences are more periodic, i.e., more beat-like.

Do Groups of Gestures Align With Spoken Boundaries?

Our initial hypothesis about perceived gesture groupings was that they would align with intonational phrases, i.e., either with Intermediate Intonational Phrases (ToBI Break Index 3) or with Full Intonational Phrases (ToBI Break Index 4). However, analysis showed that this was not reliably the case. Only 224 of 431 PGGs (51.9%) fell within a single Full Intonational Phrase, so that many PGGs appeared to extend across more than one of these prosodic constituents. This result suggested that it would be useful to extend the analysis to higher-level constituents, which might be revealed by the Extended RPT labels. Results from this analysis will be presented in two sections, addressing (1) results from the E-RPT labeling suggesting that this method captures aspects of higher-level prosodic constituent structure, and (2) results from analysis of the gestures within such higher-level constituents, suggesting that gesture sequences within those constituents tend to share kinematics to a substantial degree.

Extended Rapid Prosodic Transcription

The expansion of Cole et al.'s (2010, 2014) method for rapid 'crowd-sourced' prosodic transcription to include marking three levels of perceived boundary was undertaken in an exploratory spirit, and the very preliminary results reported here must be taken as no more than suggestive. Nevertheless, they are thought-provoking, and so we include them here.

In this preliminary study, eight participants who were not experienced prosody labelers listened to the first 2 min 15 s of the London sample, and marked three levels of perceived boundary strength by inserting one, two or three forward slashes between pairs of words where they heard these boundaries. The number of

slashes inserted by all eight participants was then totaled for each location where any participant inserted a boundary marker. Thus the highest number of boundary markers that was possible at any location was 3×8 or 24. **Figure 3** shows the total number of boundary markers inserted between a pair of words, as a function of an acoustic measure of the signal: the duration of the silence between those two words.

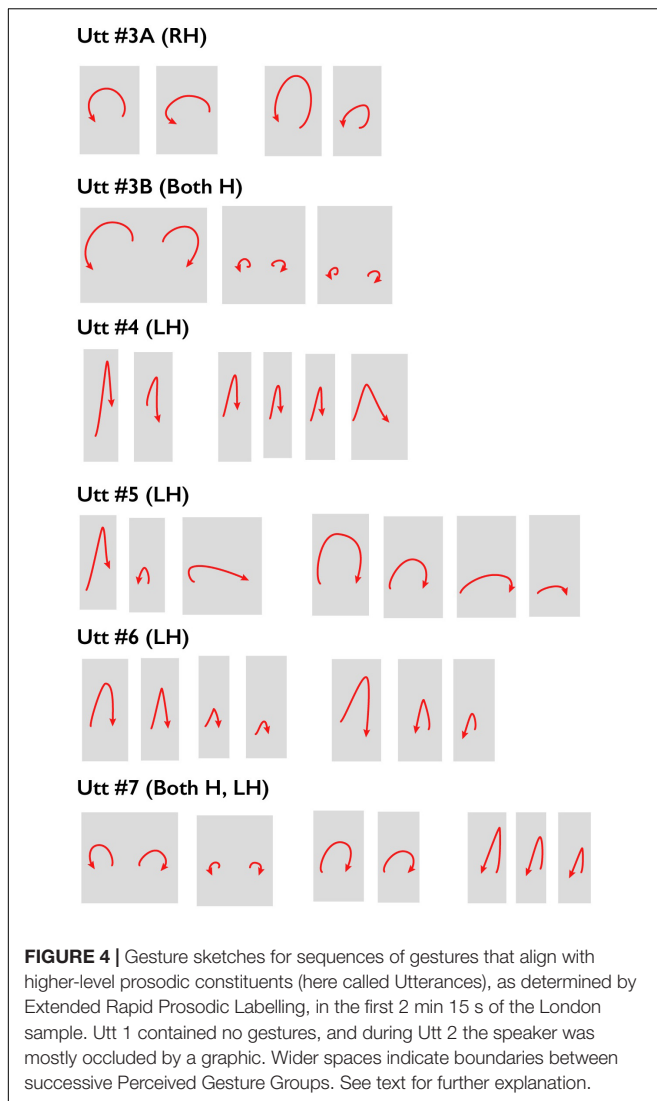
It appears that there is a reliable increase in the likelihood that the listener will perceive a boundary between two words, as a function of duration of the silence between those two words, so that the longer the silence, the more likely a listener is to insert a higher-level boundary. This observation is consistent with two inferences: that speakers organize their intonational phrases into higher level prosodic constituents, and that silence duration may be a reliable marker of these constituent boundaries. In an earlier study of possible groupings of intonational phrases into larger constituents, Wightman et al. (1992) also found hints of such a relationship between perceived higher-level constituent boundaries and silence duration. However, like the current findings, their study contained only a few such boundaries, so that the generality and reliability of the observation remains to be established by future studies.

Similarity of SDGs Within Higher-Level Constituents Identified on the Basis of E-RPT Judgments

These preliminary results suggest that the listeners' judgments reflect the silence-duration marker cue to higher-level constituents (other cues may of course also be contributing to the perception of these higher-level constituent boundaries), and they reflect a certain amount of agreement about the location of those constituents. On the assumption that this is the case, we adopted an arbitrary criterion of 15 or more boundary markers inserted by the annotators as an indicator of a higher-level grouping of individual utterances. This resulted in the identification of 8 higher-level constituents in this 2-min 15-s sample, compared to 66 Full Intonational Phrases and 100 Intermediate Intonational Phrases. Gestural sketches for the gestural accompaniments of 6 of the resulting 8 higher-level constituents are shown in **Figure 4**, where they are designated as Utterances. Visual inspection of these sketches suggests that, within a constituent defined in this way, the trajectory shape and handedness (right hand, left hand, or two hands) of successive gestures are quite similar, and that these characteristics differ from one such higher-level constituent to the next. Spacings between the sketches reflect somewhat smaller constituents defined by fewer than 15 E-RPT markings that group together smaller ToBI-labelled Full Intonational Phrases. Thus these preliminary data raise the possibility that closely related sequences of gestures are planned to occur within higher-level prosodic constituents.

Do Non-referential Gestures Form a Unified Class?

The question of how to characterize Non-Referential co-speech gestures is an important one, because the convention of referring to them as 'beats' (or sometimes 'batons') makes it easy to assume that they form a homogeneous set. But a careful reading of the



literature soon reveals that this is not the case. McNeill (1992), for example, distinguishes beats that are simple in-out or up-down 'flicks' of the hand or finger, and occur at specific points in a narrative, from other gestures that beat out the rhythm of the speech they accompany. Other researchers have also wrestled with the question of how to define and detect 'beats,' but from our point of view, a particularly interesting question concerns the ways in which a co-speech gesture can be seen as prosodic. That is, in what ways do gestures align with the prosodic prominences and constituents of the speech they accompany; in what ways do they have their own prominences and grouping structure; and in what ways do these two sets of prosodic behaviors align in time and in communicative function.

In a preliminary attempt to address these questions, we developed a system for labeling the 'beat-like-ness' of a sequence of Stroke-Defined Gestures within a PGG. The definition of this characteristic was somewhat informal, and relates to whether the sequence of movements appears to be beating out a rhythm or not. Our first attempt used a simple binary decision: is this group

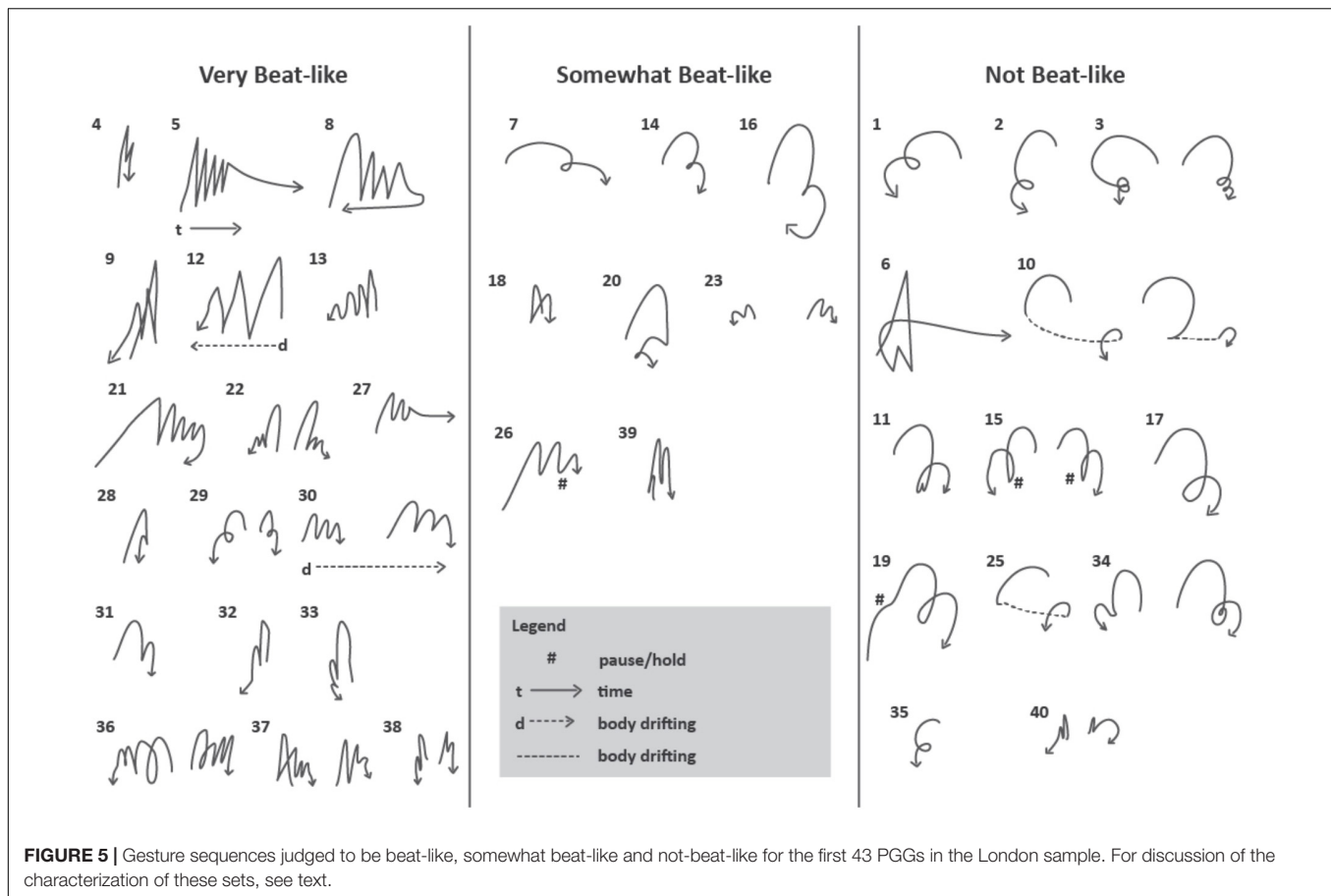
of Stroke-Defined Gestures beat-like or not, but it soon became clear that a more nuanced system was needed. We settled on a three-level categorization: beat-like, somewhat beat-like and not beat-like. The middle category, somewhat beat-like, included sequences for which some of the strokes were perceived as beat-like and others were not. (The second author, who carried out this exploratory work, would like to try a 5-level system in the future.) Results of this annotation showed that 138 (32%) of the 431 PGGs contained gesture sequences that were perceived as beat-like. 119 were labeled as somewhat beat-like, and 174 as not beat-like. A gesture sketch summary for the first 41 PGGs in the London sample is shown **Figure 5**. It appears that gestures with a straight trajectory, performed in an up-and-down vertical dimension, are more likely to be perceived as beat-like, while those with a curved trajectory are less so. A second constraint appears to be temporal: strokes of gestures judged to be beat-like occurred in quicker succession than those judged not to be beat-like. For example, the mean inter-stroke interval, measured from the end of one stroke to the beginning of the next within a PGG, was 870 ms for sequences labeled as beat-like, 992 ms for sequences labeled as somewhat beat-like, and 1,119 ms for sequences labeled as not beat-like (excluding Perceived-Gesture-Group-final tokens, for which the interval to the end of the next stroke after the PGG boundary could be very long and variable).

This result provides an initial step in the direction of distinguishing the set of Non-Referential gestures that are perceived to have a strongly rhythmic beat-like character from those with different timing characteristics. Additional work will be needed to sort out the range of possibilities for characterizing different types of Non-Referential gestures, and the ways in which both Non-Referential and Referential gestures may have different timing relationships both with other gestures and with the speech they accompany.

This discussion highlights an additional issue of some importance, which is the question of whether the common practice of designating a co-speech gesture as a member of one or another mutually exclusive category, such as 'beat-like' or 'iconic,' could be usefully supplemented by a dimension-based system, in which each co-speech gesture is annotated for all of the characteristics that it exhibits. This would permit, for example, a sequence of iconic gestures to be labeled as beatlike, if it struck the viewer as beating out the rhythm of the speech. In his article for the Cambridge Encyclopedia of Linguistic Sciences, McNeill (2006) points out the advantages of such a dimension-based approach to co-speech gesture analysis:

"The essential clue that these are dimensions and not categories is that we often find iconicity, metaphoricity, deixis and other features mixing in the same gesture. Beats often combine with pointing, and many iconic gestures are also deictic. . . A practical result of dimensionalizing is improvement in gesture coding, because it is no longer necessary to make forced decisions to fit each gesture occurrence into a single box."

Recently, Prieto et al. (2018) have discussed the multi-dimensional characteristics of beats in just these terms, and have proposed a labeling system that has many of these characteristics.



This approach may be particularly useful in the analysis of gestures which are not referential in any obvious way, but for which it is possible to imagine a metaphoric component. For example, if a speaker saying 'And thus it came to pass...' accompanies this spoken word sequence with a horizontal back-and-forth bimanual gesture with flat hands palm downward, as if smoothing a tablecloth, is that a metaphoric gesture that uses the indication of a flat smooth surface (or perhaps the act of smoothing) to stand for the concrete sequence of events to be described? This category of gesture is particularly interesting, because it encompasses gestures which bear an abstract relationship to the meaning of the speech. It sometimes seems as if almost any gestural movement can be thought of as having a metaphoric component, even though it is often difficult to put into words exactly what the potential metaphor is conveying. In a system where the degree of 'metaphoricity' could be ranked, or metaphoricity could be combined with other dimensions such as rhythmicity, such problems might be less vexing.

We note in passing that this sample includes very few of the hand or finger 'flicks' identified by McNeill (1992): only 23 examples of in-out flicks were identified, i.e., 1.7% of the total number of Stroke Defined Gestures. It is possible that this is due to the fact that this speaker was standing up behind a podium, with no place to rest his arms and hands, in contrast to speakers

who produce a narrative while sitting in a chair with arms where they often rest their own arms and hands. This might make a finger-flick more comfortable. Another possibility is that in this sample, the function of a finger flick is served by a larger vertical movement of the entire arm and hand. However, that seems unlikely since such vertical movements are quite common in this sample, and often give the impression of being comprised of a preparation and a stroke, rather than of a bi-phasic in-out or up-down 'flick.'

DISCUSSION

The observation that most of the co-speech gestures in this sample are judged to be Non-Referential has provided an opportunity to examine some of the characteristics of this type of gesture. The preliminary results presented here suggest that the strokes of these Non-Referential gestures align with prominent (i.e., pitch-accented) syllables in the speech they accompany, as has been reported for small samples of Referential gestures and for larger undifferentiated samples. In addition, preliminary observation raises the possibility that they group into constituents that align with higher-level prosodic constituents. This is consistent with the possibility of a parallel signaling of the organization of gestural and speech constituents at a level higher than the individual intonational phrase or even utterance,

as proposed in, e.g., Kendon's hierarchy of prosodic/gestural constituents (Kendon, 1972, 1980, 2004) and suggested by McNeill's 'cohesive' gestures (McNeill, 1992). The observations reported here also suggest that Non-Referential co-speech gestures are not a homogeneous class, either kinematically or functionally, but instead may contain a wide variety of forms and serve a wide range of communicative ends. This raises the question of how the process of generating co-speech gestures can be integrated into current models of speech production planning. We will discuss each of these points in turn.

The Alignment of Non-referential Gestures With Phrase-Level Prosodic Prominences

The question of how speakers determine the alignment of speech and co-speech gesture raises the methodological question of how best to define and study the alignment of spoken prosody with co-speech gestural events. With respect to the methodological question, a range of criteria for accent-gesture alignment have been used, from the strict temporal overlap of accented syllable with gestural stroke employed in this study, to a more expansive criterion of the two events being within a pre-defined number of milliseconds (e.g., Loehr, 2004), and from the alignment of temporal intervals (strokes with pitch-accented syllables) to the alignment of precise time points (F0 maxima, gesture apices). Brentari et al. (2013) suggest an interesting hierarchy of alignments, ranging from exact correspondence of the two time intervals, to overlap of the accented syllable with at least part of the stroke, to overlap with at least part of the entire gesture (including any preparation, hold, and recovery phases). They note that listeners can form an impression of which word a gesture is associated with, even when there is no direct temporal alignment. The question of what 'counts' as alignment/association between a spoken word and a gesture is clearly in need of investigation. Studies by Renwick et al. (2004), Yasinnik et al. (2004), and Shattuck-Hufnagel et al. (2007) measured alignment of manual strokes that had short sharp end points (resulting in a clear rather than a blurry video frame), which they called 'hits,' in a different set of academic lecture videos. Results showed that these end points occurred reliably toward the end of or just after a spoken accented syllable. Other investigations have focused on the alignment of the *onset* of a gesture or the *apex* of a stroke with an aspect of the speech. A model of speech production planning that includes gestural planning will need to specify which part of the gesture is planned to align with which part of the speech, in cases where that relationship is shown to be systematic.

Beyond the question of precisely how strokes and accented syllables are aligned, a larger question concerns which accented words and syllables are accompanied by co-speech gestures and which ones are not. For Referential gestures, earlier observations showed that the stroke is likely to overlap with a phrase-level prominence/pitch accent (e.g., Kendon, 2004). The finding that, in the sample of largely Non-Referential gestures examined here, 83% of the stroke intervals overlap in time at least partially with a pitch-accented syllable interval also reveals that 17% did not. Why are some strokes produced in non-accented regions of

the speech? In addition, many pitch-accented syllables are not accompanied by a co-speech gesture. What determines which accents are aligned with strokes and which accents are not? This question awaits further study.

The Alignment of Co-speech Gestures With Higher-Level Prosodic Constituents

The preliminary observation that perceived higher-level prosodic constituents in the speech may overlap with sequences of kinematically similar Non-Referential gestures raises the question of the precise nature of these constituents. Kendon (1980) notes that, in his observations, prosodic Tone Groups are combined into higher-level Locutions (said to generally comprise a complete sentence), which are in turn combined into Locution Clusters within a Discourse or conversational turn. These higher levels of constituent structure do not figure prominently in the Autosegmental-Metrical model of prosodic structure which was initially adopted for this study, in part because they have not been observed to have clear intonational markers. The Extended Rapid Prosodic Transcription method may prove useful in identifying acoustic cues that are specific markers for these higher level structures, like the duration-of-silence correlate discussed above. As Kendon proposed, some of the cues to these higher-order structures may be found in the gestural domain, in the sense that sequences of similar gestures may align with such constituents, so that a change in a gestural dimension might mark the start of a new constituent. If so, it will be consistent with the view that models of human speech production planning (and speech perception) must expand to accommodate the ways in which speakers insert this kind of information into the visual signal.

An interesting aspect of these preliminary observations is that they suggest subgroupings below the level of boundary corresponding to the arbitrary criterion adopted here (15 boundary markers). For example, in the set of gestures within the first of the higher-level spoken prosodic constituents shown in **Figure 4**, there appears to be a shift in gesture kinematics halfway through the constituent (i.e., between Utterance 3A and Utterance 3B); this corresponds to a location where the annotators inserted 14 boundary markers, a value which is just under our arbitrary threshold. The suggestion of a lower-level constituent boundary in the Extended Rapid Prosodic Transition data at that location is consistent with the change in gesture kinematics at that point. Similarly, in a later part of this sample, where listeners annotated the word sequence 'and only nearly lost it, once' as a single higher level constituent, but also indicated a smaller perceived boundary after 'lost it,' the trajectory shape of the Stroke-Defined Gesture produced with 'once' is different from that of the SDGs produced with the preceding word sequence 'and only nearly lost it.' Such observations support the possibility that, like prosodic constituents, gesture sequences are hierarchically organized.

Finally, the question of what signals the grouping of a sequence of gestures into a constituent has been only tangentially addressed in this paper. In the larger ongoing project of which this study is a part, the visual-only annotation of gesture groups employed to identify PGGs is supplemented with gesture phase

labeling. This will enable testing the hypothesis advanced in Kendon (2004) that groups of gestures (Gesture Units, in his terminology) are marked by a recovery phase at the end of the group-final gesture, in this sample of academic-lecture-style speech that contains mostly Non-Referential gestures. Moreover, combining video-only labeling of PGGs (which focuses on the physical characteristics and timing of the gestural movements), with sound-only labeling of the spoken prosodic constituents, allows the investigation of timing and grouping alignments between the two streams of behavior. In the end, however, by combining these separate annotation approaches with listening and looking at the same time, it may be possible to determine how the semantic, syntactic, prosodic, and gestural structures of an utterance combine to form an effective act of communication. Modeling that process will require a collaborative effort which, it is hoped, this report may help to inspire.

Integrating Gesture Production Planning Into Current Speech Production Models

The results described in this paper offer support for the hypothesis put forward over the years by Kendon and McNeill and others, that the gestures that accompany a spoken utterance are an integral part of the communication signal, and thus that the planning process for producing a spoken utterance must include the planning of co-speech gestures. In particular, taken together with other results in the literature, these findings suggest a tight temporal coordination between the prosodic structure (i.e., the grouping and prominence structure) of a spoken utterance and the prosodic structure of the gestures that accompany it. In this way they are also consistent with the hypothesis proposed by Keating and Shattuck-Hufnagel (2002), that the planning frame for a spoken utterance is a prosodic planning frame. Keating and Shattuck-Hufnagel propose a 'Prosody First' model of the phonological encoding process in speech production planning. In that model, a representation of the phrase-level prosody of an utterance is computed as an abstract structure, simple at the beginning of the planning process but gaining complexity as the phonological elements of the planned utterance are inserted into its sequentially and hierarchically organized slots. On this view, the prosodic structure of an utterance provides the representational 'spine' that governs the serial ordering of lexical elements and their sub-constituents, the integration of multiple factors involving the surface timing/duration patterns of the speech signal, and the computation of surface timing patterns.

For a more comprehensive view of speech production planning that begins with the earliest formation of the intended message, one can turn to the model proposed by Levelt (1989) and implemented by Levelt et al. (1999). In this model the initial formulation of a message takes place in terms of a cognitive representation of meaning that is pre-linguistic. It may be this very early representation that guides the subsequent formation of both the spoken and the gestural realizations of the utterance. Kendon (1980) suggests this when he notes that

"we may mention the views of Chafe (1970) who has argued explicitly for the position that the process of utterance

generation proceeds through a series of steps starting with the organization of semantic structures. The work on gesticulation we have reviewed here would suggest that this earliest stage in the process of utterance formation has, or can have, direct expression in gesticular action." (p. 224)

Jannedy and Mendoza-Denton (2005) expressed a related idea in their report of a study of co-speech gesture timing and function in a sample of highly emotional (and presumably not highly rehearsed) political speech:

"We found that speech and gesturing are two different channels/modes of information transfer which allow for different content to be transmitted. If we assume the validity of Bolinger's (1986) claim that "gesture and speech stem from the same semantic intent [...]," then we commit ourselves to the notion that some degree of pre-planning is involved in generating not only speech output but also gestural output in order to convey information on different planes. How information is structured and divided up across the two channels is not understood at this point. From our data it appears that complementary and contextual information is transmitted via gestures while concrete assertions are made explicit via speech. We also do not know what constraints exist in (pre-)planning complex gestures that we know are time-aligned with linguistic structure in the final output." (Jannedy and Mendoza-Denton, 2005, p. 233).

CONCLUSION

The results described in this paper offer support for the hypothesis put forward over the years by Kendon and McNeill and their colleagues, that the gestures that accompany a spoken utterance are an integral part of the communication signal, and thus that the planning process for producing a spoken utterance must include the planning of co-speech gestures. In particular, taken together with other results in the literature, these findings suggest a tight temporal coordination between the prosodic structure of a spoken utterance and the prosodic structure of the gestures that accompany it. In this way they are also consistent with the hypothesis proposed by Keating and Shattuck-Hufnagel (2002), that the planning frame for a spoken utterance is a prosodic planning frame. On this view, the prosodic structure of an utterance provides the representational 'spine' that governs not only the serial ordering of lexical elements and their sub-constituents, the integration of multiple factors involving the surface timing/duration patterns of the signal, and the computation of surface timing patterns, but also, potentially, the integration of auditory with visual aspects of the speech act. On this hypothesis, the prosodic planning frame governs the timing of occurrence and the duration of various components of both the spoken and the gestural aspects of the communicative act (Shattuck-Hufnagel et al., 2016).

It must be emphasized that the results reported here are drawn from a single speaker, producing speech in a particular context and style, and it is not yet clear how far they will generalize.

Moreover, although some of the results reported here are based on large numbers of manual gestures and spoken prosodic events, others are based on very small numbers of observations and are thus highly preliminary. However, in concert with other observations in the literature, the results reported here open the door to a number of lines of study. These include the investigation of the cues to higher-level prosodic constituents that group spoken intonational phrases together, and of the patterns in the use of individual cues to prosodic constituents in both the spoken and the gestural domains that may vary across speakers, listeners, learners, and users of different languages. It appears that the study of how co-speech gestures and speech interact in communication systems is poised on the threshold of some very interesting discoveries which will enlarge and enhance our ability to build models of the speech production planning and speech perception processes.

ETHICS STATEMENT

This work was carried out under approval of MIT's Committee on the Use of Human Experimental Subjects (COUHES).

REFERENCES

- Beckman, M., and Pierrehumbert, J. (1986). "Intonational structure in Japanese and English," in *Phonology Yearbook*, eds C. J. Ewen and J. M. Anderson (Cambridge: Cambridge University Press), 15–70.
- Bolinger, D. (1986). Intonation and gesture. *Am. Speech* 52, 156–174.
- Brentari, D., Marotta, G., Margherita, I., and Ott, A. (2013). *The Interaction of Pitch Accent and Gesture Production in Italian and English*. Chicago, IL: University of Chicago, 79–97.
- Byrd, D., Krivokapic, J., and Lee, S. (2006). How far, how long: on the temporal scope of prosodic boundary effects. *J. Acoust. Soc. Am.* 120, 1589–1599. doi: 10.1121/1.2217135
- Chafe, W. (1970). *Meaning and the Structure of Language*. Chicago: Chicago University Press.
- Cole, J., Mahrt, T., and Hualde, J. I. (2014). "Listening for sound, listening for meaning: task effects on prosodic transcription," in *Proceedings of the Speech Prosody*, Dublin, 7.
- Cole, J., Mo, Y., and Hasagawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *J. Lab. Phonol.* 1, 425–452. doi: 10.1515/labphon.2010.022
- Efron, D. (1941/1972). *Gesture, Race and Culture*. The Hague: Mouton.
- Esteve-Gibert, N., Borrás-Comes, J., Asor, E., Swerts, M., and Prieto, P. (2017). The timing of head movements: the role of prosodic heads and edges. *J. Acoust. Soc. Am.* 141, 4727–4739. doi: 10.1121/1.4986649
- Esteve-Gibert, N., and Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *J. Speech Lang. Hear. Res.* 56, 850–864. doi: 10.1044/1092-4388(2012/12-0049)
- Ferre, G. (2005). "Gesture, intonation and the pragmatic structure of narratives in British English conversations," in *York Papers in Linguistics*. York: Department of Language and Linguistic Science, eds G. Hicks and G. Walker (Toronto, ON: University of York), 55–90.
- Ferre, G. (2010). "Timing relationships between speech and co-verbal gestures in spontaneous French," in *Proceedings of the Language Resources and Evaluation, Workshop on Multimodal Corpora*, Malta, 86–91.
- Jannedy, S., and Mendoza-Denton, N. (2005). Structuring information through gesture and intonation. *Interdiscip. Stud. Inf. Struct.* 3, 199–244.

AUTHOR CONTRIBUTIONS

SS-H generated the hypotheses to be tested, developed the Extended RPT method, labeled the prosody, participated in the development of the gesture labeling methods and analysis of the data, and wrote most of the text. AR labeled the gestures, developed the gesture sketch method, participated in the development of the gesture labeling methods and data analysis, and generated the figures.

FUNDING

This work was supported in part by NSF grants to SS-H (BCS 1023596 & 1205402); by the MIT Speech Communication Group's North Fund and Klatt Fund; and by MIT's Undergraduate Research Opportunities Program.

ACKNOWLEDGMENTS

We gratefully acknowledge the contributions of the members of the MIT Freshman Seminar in Prosody (Fall 2017).

- Keating, P., Baroni, M., Mattys, S., Scarborough, R., Alwan, A., Auer, E. T., et al. (2003). "Optical phonetics and visual perception of lexical and phrasal stress in English," in *Proceedings of the ICPHS 2003 conference*, (Spain: Barcelona), 2071–2074.
- Keating, P. A., and Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Work. Paper. Phonet.* 101, 112–115.
- Kendon, A. (1972). "Some relationships between body motion and speech: an analysis of an example," in *Studies in Dyadic Communication*, eds A. Siegman and B. Pope (Elmsford, NY: Pergamon), 177–210. doi: 10.1016/B978-0-08-015867-9.50013-7
- Kendon, A. (1980). "Gesticulation and speech: two aspects of the process of utterance," in *The Role of Nonverbal Communication*, ed. M. R. Kay (Berlin: De Gruyter Mouton), 207–227.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572
- Krahmer, E., Ruttkay, Z., Swerts, M., and Wesselink, W. (2002). "Pitch, eyebrows and the perception of focus," in *Proceedings of the Speech Prosody 2002*, (Aix-en-Provence), 443–446.
- Krivokapic, J. (2014). Gestural coordination at prosody boundaries and its role for prosodic structure and speech planning processes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130397. doi: 10.1098/rstb.2013.0397
- Krivokapic, J., Tiede, M., and Tyrone, M. E. (2015). "A kinematic analysis of prosodic structure in speech and manual gestures," in *Proceedings of the 18th International Congress of Phonetic Sciences*, (Glasgow: the University of Glasgow).
- Levelt, W. J. M. (1989). *Speaking: from Intention to Articulation*. Cambridge: MIT Press.
- Levelt, W. J. M., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75. doi: 10.1017/S0140525X99001776
- Liberman, M., and Prince, A. (1977). On stress and linguistic rhythm. *Linguist. Inq.* 8, 249–336.
- Loehr, D. (2004). *Gesture and Intonation*. Ph.D. thesis, Georgetown University, Washington, DC.
- Loehr, D. (2012). Temporal, structural and pragmatic synchrony between intonation and gesture. *J. Lab. Phonol.* 3, 71–89. doi: 10.1515/lp-2012-0006
- McClave, E. (2000). Linguistic functions of head movements in the context of speech. *J. Pragmat.* 32, 855–878. doi: 10.1016/S0378-2166(99)00079-X

- McClave, E. Z. (1994). Gestural beats: the rhythm hypothesis. *J. Psychol. Res.* 23, 45–65. doi: 10.1007/BF02143175
- McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.
- McNeill, D. (2006). “Gesture: a psycholinguistic approach,” in *The Encyclopedia of Language and Linguistics* (Amsterdam & Boston: Elsevier), 58–66.
- McNeill, D., and Levy, E. (1982). “Conceptual representations in language and gesture,” in *Speech, Place and Action*, eds R. J. Jarvella and W. Klein (Chichester: John Wiley and Sons), 271–296.
- Myers, J. (2012). *The Behavior of Hand and Facial Gestures at Pauses in Speech*. Ph.D. thesis, Yale University, New Haven, CT.
- Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.
- Pierce, C. S. (1960). “Collected papers of Charles Sanders Pierce (1931–1958),” in *Volume 1: Principles of Philosophy. Volume 2: Elements of Logic*, eds C. Hartshorne and P. Weiss (Cambridge, MA: The Belknap of Harvard University Press).
- Prieto, P., Cravotta, A., Kushch, O., Rohrer, P. L., and Vila-Gimenez, I. (2018). “Deconstructing beats: A labelling proposal,” in *Proceedings of the Speech Prosody 2018*, Poznan.
- Renwick, M., Shattuck-Hufnagel, S., and Yasinnik, Y. (2004). The timing of speech-accompanying gestures with respect to prosody (Abstract). *J. Acoust. Soc.* 115:2397. doi: 10.1121/1.4780717
- Selkirk, E. O. (1984). *Phonology and Syntax: The Relation Between Sound and Structure*. London: MIT Press.
- Shattuck-Hufnagel, S., and Ren, A. (2012). “Preliminaries to a Kinematics of Gestural Accents,” in *Presented at the International Society for Gesture Studies 6*, San Diego.
- Shattuck-Hufnagel, S., Ren, A., Mathew, M., Yuen, I., and Demuth, K. (2016). Non-referential gestures in adult and child speech: are they prosodic? *Proc. Speech Prosody 2016*, 836–839. doi: 10.21437/SpeechProsody.2016-171
- Shattuck-Hufnagel, S., Ren, A., and Tauscher, E. (2010). “Are torso movements during speech timed with intonational phrases?,” in *Proceedings of the Speech Prosody 2010*, Chicago, IL, 974.
- Shattuck-Hufnagel, S., and Turk, A. (1996). A prosodic tutorial for investigators of auditory speech processing. *J. Psychol. Res.* 25, 193–247. doi: 10.1007/BF01708572
- Shattuck-Hufnagel, S., Yasinnik, Y., Veilleux, N., and Renwick, M. (2007). “A method for studying the time alignment of gestures and prosody in American English: ‘Hits’ and pitch accents in academic-lecture-style speech,” in *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, eds A. Esposito, M. Bratanic, E. Keller, and M. Marinaro (Brussels: NATO), 34–44.
- Swerts, M., and Krahmer, E. (2010). Visual prosody of newsreaders: effects of information structure, emotional content and intended audience on facial expressions. *J. Phon.* 38, 197–206. doi: 10.1016/j.wocn.2009.10.002
- Turk, A., and Shattuck-Hufnagel, S. (2007). Phrase-final lengthening in American English. *J. Phon.* 35, 445–472. doi: 10.1016/j.wocn.2006.12.001
- Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction. *Speech Commun.* 57, 209–232. doi: 10.1016/j.specom.2013.09.008
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* 91, 1707–1717. doi: 10.1121/1.402450
- Yasinnik, Y., Renwick, M., and Shattuck-Hufnagel, S. (2004). “The timing of speech-accompanying gestures with respect to prosody,” in *Proceedings of the From Sound to Sense: 50+ years of Discoveries in Speech Communication*, (Cambridge MA: MIT), C97–C102.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Shattuck-Hufnagel and Ren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Prosody in the Auditory and Visual Domains: A Developmental Perspective

Núria Esteve-Gibert^{1*} and Bahia Guellai^{2*}

¹ Departament de Llengües i Literatures Modernes i d'Estudis Anglesos, Universitat de Barcelona (UB), Barcelona, Spain,

² Laboratoire Ethologie, Cognition, Développement, Université Paris Nanterre, Nanterre, France

OPEN ACCESS

Edited by:

Marianne Gullberg,
Lund University, Sweden

Reviewed by:

Mili Mathew,
St. Cloud State University,
United States

Francesca Marina Bosco,
Università degli Studi di Torino, Italy

*Correspondence:

Núria Esteve-Gibert
nuria0esteve0gibert@gmail.com
Bahia Guellai
bahia.guellai@gmail.com

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 15 December 2017

Accepted: 27 February 2018

Published: 19 March 2018

Citation:

Esteve-Gibert N and Guellai B
(2018) Prosody in the Auditory
and Visual Domains: A Developmental
Perspective. *Front. Psychol.* 9:338.
doi: 10.3389/fpsyg.2018.00338

The development of body movements such as hand or head gestures, or facial expressions, seems to go hand-in-hand with the development of speech abilities. We know that very young infants rely on the movements of their caregivers' mouth to segment the speech stream, that infants' canonical babbling is temporally related to rhythmic hand movements, that narrative abilities emerge at a similar time in speech and gestures, and that children make use of both modalities to access complex pragmatic intentions. Prosody has emerged as a key linguistic component in this speech-gesture relationship, yet its exact role in the development of multimodal communication is still not well understood. For example, it is not clear what the relative weights of speech prosody and body gestures are in language acquisition, or whether both modalities develop at the same time or whether one modality needs to be in place for the other to emerge. The present paper reviews existing literature on the interactions between speech prosody and body movements from a developmental perspective in order to shed some light on these issues.

Keywords: speech, gestures, prosody, development, multimodality

INTRODUCTION

Human language is an interesting input as it can be perceived through both ears and eyes. For example, adults' comprehension of speech in noisy and quiet environments is enhanced when they have access to the visual cues conveyed by the speaker's face (Sumbly and Pollack, 1954). In face-to-face interactions, the whole body is involved and may serve informative purposes (Kelly and Barr, 1999 for a review; Kendon, 2004). People around the world produce spontaneous gestures while talking. These gestures accompanying speech, called 'co-speech gestures,' are so connected with speech that people use their hands even when nobody sees them (Corballis, 2002), and congenitally blind people gesture when interacting with each other (Iverson and Goldin-Meadow, 1998 and Goldin-Meadow, 1998). Gestures can be defined on the basis of the articulator that is being used to produce them (the head, as in head nods or head tilts; the hand, as in manual pointing, manual beats or iconic gestures; the face, as in oral gestures or in facial expressions such as eyebrow movements), on the basis of whether or not they are accompanied by speech (co-speech gestures), or based on whether the gesture movement is continuous or discrete (see Wagner et al., 2014 for a review). Another order of things is the function for which they are used in language and communication. Gestures can serve a deictic or highlighting function, they can depict and represent semantic meanings, and they can structure information in the discourse and be an indicator of pragmatic

implicatures to be driven for a successful communication to take place. Because all these levels have parallels with the prosodic properties of speech, these gestures are also called visual correlates of prosody.

It is now clear that co-speech gestures fulfill multiple cognitive functions. Some studies focused on speaker-directed functions suggesting that gestures may ease the speaker's cognitive load (Cook and Goldin-Meadow, 2006; Chu and Kita, 2011), promote learning (Ping and Goldin-Meadow, 2010), help in the conceptual planning of information and discourse (Alibali et al., 2000; Cutica and Bucciarelli, 2008), and facilitate lexical access (Rauscher et al., 1996; Alibali et al., 2000). Others stress that gestures enhance the transfer of information by providing it cross-modally, thereby facilitating uptake for addressees (De Ruiter et al., 2012; Guellai et al., 2014). These proposals account for the adults' use of co-speech gestures and focus on gestures with a referential value in communication (deictic and iconic hand movements). Yet, they are less effective for explaining developmental patterns as well as the role of gestures with a non-referential value in communication (such as facial expressions and rhythmic 'beats').

In the following sections we propose to explore the developmental links between speech and body movements (i.e., hand and head gestures, and facial expressions), focusing on one specific linguistic aspect, namely prosody. Prosodic properties of speech encode prominence, phrasal organization, speech act types, emotions, attitudes, and beliefs (e.g., Pierrehumbert and Hirschberg, 1990; Ladd, 1996; Byrd and Saltzman, 2003; Jun, 2005). There is a growing body of research showing that prosody is not only expressed through the tonal and temporal properties of speech, but also by means of body movements produced with the hand, head, or face (e.g., Krahmer and Swerts, 2007; Cvejic et al., 2012; Guellai et al., 2014). The speech and gesture dimensions of prosody are found to be tightly intertwined at the temporal, semantic, and pragmatic levels, and this is true not only in adult speech but also in language development.

Speakers' body movements are temporally coordinated with the prosodic structure in speech, pitch accents and boundary tones serving as anchoring points for prominent phases in body movements (Hadar et al., 1983; De Ruiter, 1998; Leonard and Cummins, 2011; Esteve-Gibert and Prieto, 2013; Ishi et al., 2014; Ambrazaitis and House, 2017; Esteve-Gibert et al., 2017a). At the semantic and pragmatic levels, prosody and gestures can both have a deictic component through which speakers highlight certain elements in speech (Levelt et al., 1985; Roustian and Dohen, 2010), they can disambiguate syntactic constituents (Guellai et al., 2014; Krivokapic et al., 2016), and mutually influence the processing of speaker's emotions, beliefs, and attitudes (Ekman, 1979; Kendon, 2004; Poggi et al., 2013). In the multimodal expression of prosody, the gesture dimension can consist of movements of the hand or head, facial expressions, or body postures. Traditionally, different types of body movements have been studied independently (for instance, facial expressions have received more attention in the literature on emotions, while hand movements have been the focus of studies on the referential value of gestures in language). In the present paper we will refer to these different types of movements as 'gestures,' as we propose

that it is more interesting to take them as a whole to have a complete picture of the speech-gesture relationship in language and communication development.

TEMPORAL ASPECTS OF THE AUDIO-VISUAL SPEECH INTEGRATION IN INFANCY

Infants need to make sense of the rich multisensory stimulations present in their everyday experiences. From the earliest stages of development, infants are found to relate phonetic information from the lips and the voice (Kuhl and Meltzoff, 1984; Aldridge et al., 1999; Patterson and Werker, 2003). In these studies, infants were presented with videos, side-by-side, of two faces articulating two vowels (i.e., /i/ vs. /a/), while hearing only one vowel (i.e., either /i/ or /a/). Infants are considered to be able to detect audio-visual congruency if they look longer at the matching stimulus. Remarkably, there is evidence that from birth, infants detect equivalent phonetic information in the lips and voice (Aldridge et al., 1999). Auditory-visual phonetic matching is also shown at 2 months (Patterson and Werker, 2003), at 4 months and a half (Patterson and Werker, 1999), and at 8 months based on the gender of the talker (Patterson and Werker, 2002). When the vowels are reduced to sine-wave analogs or simple tones, infants do not detect the congruent video anymore (Kuhl et al., 1991). Taken together, these studies, focusing on perioral and facial cues, suggest that infants already have the primitives of lip reading for single speech sounds.

On the production side, newborns bring their hands and objects to their mouth, and explore them orally, these behaviors being considered to be the earliest signs of the oral-manual link in language development (Iverson and Thelen, 1999). Around 6–7 months of age infants start to babble, a rhythmic close-open movement of the jaw that results in the production of syllables (Oller, 2000; Vihman et al., 2009). At the same age infants start producing rhythmic arm movements that are temporally aligned with the vocal babbling (Ejiri, 1998; Iverson and Fagan, 2004). Interestingly, the acoustic quality of the infants' babbles improves when infants combine these vocalizations with rhythmic arm movements, as syllables become shorter and display shorter formant-frequency transitions (Ejiri and Masataka, 2001).

The time-aligned coordination of gesture and speech is also present at later stages of language development. At the onset of word production infants start combining vocalizations with pointing gestures signaling referents in space, and these gestural and speech dimensions are timely aligned in an adult-like way: the accented syllable in speech coincides with the apex of the pointing gesture (Butcher and Goldin-Meadow, 2000; Esteve-Gibert and Prieto, 2014). Later, at 4–5 years of age we observe the emergence of bi-phasic body movements that have no referential meaning and that are timed with pitch accents that children use to emphasize specific information in the sentence (Nicoladis et al., 1999; Capone and McGregor, 2004; Esteve-Gibert et al., 2017b; Mathew et al., 2017). These movements are typically produced with the hand, arm, or head, and are called beats in the gesture literature (Kendon, 2004; McNeill, 2005; Wagner et al., 2014).

Beats provide clear evidence of the rhythmic entrainment between the acoustic and visual dimensions of language, because speakers are found to necessarily modify the acoustic properties of speech when they produce these body movements (Krahmer and Swerts, 2007). Thus, prosodic structure seems to be observed at the speech and at the gestural levels, both dimensions being temporally aligned in a precise way from early stages of language development.

IMPLICATIONS OF THE AUDIO-VISUAL INTEGRATION FOR WORD LEARNING

When addressing infants, adults usually use a speech register which is commonly called Infant-Directed Speech (IDS). This speech register has been the focus of numerous studies as it presents particularities in the auditory domain. It is characterized by slower speech rate and exaggerated pitch excursions compared to Adult-Directed Speech (ADS) (e.g., Fernald and Simon, 1984; Grieser and Kuhl, 1988; Fisher and Tokura, 1995). Vowel and consonant contrasts are more clearly produced in IDS, and this acoustic difference helps infants to build their phoneme inventories (Kuhl et al., 1991; Werker et al., 2007; Cristia, 2011). Also, the slower speaking rate and vowel properties help 21-month-olds learn and remember new words better (Song et al., 2010; Ma et al., 2011).

It has also been observed that IDS is associated with exaggerated facial cues: when addressing infants, caregivers usually exaggerate facial expressions and articulatory lip gestures for corner vowels (Chong et al., 2003; Green et al., 2010). It has been argued that visual IDS attracts infants' attention to the speaker and helps them to parse the speech stream (Kitamura and Burnham, 2003). Some authors have examined sensitivity to the temporal synchrony of visual prosody using continuous IDS (Blossom and Morgan, 2006). They found that infants aged 10–11 months use visual prosody to extract information about the structure of language as they matched synchronous faces and voices. More recently, it has been shown that 8-month-old infants reliably detect congruence between matching auditory and visual displays of a talking face based on prosodic motion (Kitamura et al., 2014), and that 9-month-olds can detect whether a manual deictic gesture is congruently aligned with the corresponding speech segment (Esteve-Gibert et al., 2015). Using an intermodal matching paradigm, Kitamura et al. (2014) presented 8-months-old infants with two visual displays of talking faces (i.e., only moving dots) and one utterance that matched one of the two facial configurations. Results showed that infants reliably detect auditory and visual congruencies in the displays. It seems that this ability emerges early in development as newborns are already able to match a facial display to the corresponding speech stream (Guellaï et al., 2016).

Another dimension of IDS is found in the body gestures of caregivers, which trigger and enhance speech processing. Indeed, caregivers accompany speech with deictic and iconic gestures when talking about objects and actions to infants (Clark and Estigarribia, 2011; Esteve-Gibert et al., 2016), and highlight referential communication by labeling objects while

moving them in synchrony with speech (Gogate et al., 2000; Jesse and Johnson, 2016). The caregivers' use of co-speech gestures seems to boost infants' receptive vocabulary and memory skills (Goodwyn et al., 2000; O'Neill et al., 2005; Zammit and Schafer, 2011; Igualada et al., 2017). Igualada et al. (2017) tested preschoolers in a word learning task in which certain words in the list were accompanied by a beat gesture, and results indicated that words co-occurring with gestures were better remembered than gesturally unmarked words.

Yet the impact of Infant-Directed Gestures (or 'gestures') on language development is an unresolved issue. Some studies have found that toddlers learn words better if adults accompany object labels with deictic and symbolic gestures, and direct their gaze toward the object (Booth et al., 2008; McGregor et al., 2009). However, other findings do not support this hypothesis, some results showing an absence or very small effect of parental use of deictic and symbolic gestures on infants' word learning abilities (Zammit and Schafer, 2011; Puccini and Liszkowski, 2012).

MULTIMODAL DEVELOPMENT OF DISCOURSE AND NARRATIVE SKILLS

An interesting aspect of prosody is that it can also convey information about syntax (Nespor and Vogel, 1986, 2007; Langus et al., 2012). For example, one can manipulate prosodic cues to influence how listeners interpret syntactically ambiguous sentences (Lehiste, 1973; Cooper and Paccia-Cooper, 1980; Price et al., 1991; Carlson et al., 2001). These effects emerge very quickly during sentence comprehension (Marslen-Wilson et al., 1992; Warren et al., 1995; Nagel et al., 1996; Kjølgaard and Speer, 1999; Weber et al., 2006). In the visual domain, the so-called beat gestures seem to be also used to process the structure of the speech signal. In languages such as Italian, English, Dutch, or Catalan, beat gestures are temporally aligned with pitch accents and boundary tones (Yasinnik et al., 2004; Krahmer and Swerts, 2007; Esteve-Gibert et al., 2017a; Krivokapic et al., 2017). Guellaï et al. (2014) showed that spontaneous gestures accompanying speech can be perceived as prosodic markers by adults. This evidence goes in the same direction as a model based on Israeli Signed Language (ISL) showing that body positions align with rhythmic manual features of the signing stream to mark prosodic boundaries (Nespor and Sandler, 1999; Sandler, 1999, 2005, 2011, 2012).

Speakers use prosodic means to emphasize new and important information in ongoing discourse, and for signaling the conceptual structure of the utterances in narrations (Swerts and Geluykens, 1994; Gussenhoven, 2004; Baumann and Grice, 2006; Ladd, 2008). Likewise, visual strategies are found to serve similar functions. Articulatory and head gestures enhance the perception of contrastive focus (Dohen and Loevenbruck, 2009; Swerts and Krahmer, 2010; Kim et al., 2014; Prieto et al., 2015), and body gestures such as eyebrow and head movements are produced less often as a marker of the theme than as a rheme marker (Ambraszaitis and House, 2017).

Children develop discourse and narrative skills relatively late. At around 5 years of age, children use adult-like discourse

markers, dependent clauses and sentential focus to narrate actions with a coherent structure, and these abilities continue to develop over the next years (Hudson and Shapiro, 1991; Berman and Slobin, 1994; Diessel and Tomasello, 2005; Kallay and Redford, 2016). The question is whether gesture and prosodic markers emerge together with the development of syntactic and lexical markers of conceptual structure. On the gesture side, at ages four to five children use beat gestures to emphasize specific information in the sentence (Nicoladis et al., 1999; Capone and McGregor, 2004; Esteve-Gibert et al., 2017b; Mathew et al., 2017). In narrations, children seem to gesture more when they produce longer sentences with more connectives (Nicoladis et al., 1999; Graziano, 2011, 2014; Colletta et al., 2014), and they use different gesture types depending on the age and the type of discourse they produce (Alamillo et al., 2013). Also, they display better narrative skills in a story retelling game if they have had access to manual beat gestures marking information focus and event boundaries (Vilà-Giménez et al., 2017). On the speech prosody side, children at age five and six are found to use the appropriate pitch accents with the right alignment to signal new information in the discourse (see Chen, 2018 for a review), and in narratives they mark event boundaries through pitch direction and linearity (Kallay and Redford, 2016). While results from the gesture literature seem to suggest that gesture marking of discourse structure is directly correlated with the development of linguistic skills, results are less conclusive from the speech prosody side. Kallay and Redford (2016) propose that the correlation between the development of linguistic skills and the development of discourse structure might occur at the level of local pitch features, while more global aspects of discourse prosody such as slope steepness, pitch resets, or pause duration might be mediated by non-linguistic factors such as breathing.

MULTIMODAL CUES IN DEVELOPING EMOTION PERCEPTION AND PRODUCTION

Perceptual skills related to emotion develop very early in infancy. It has been found that 5-month-old infants are able to distinguish between two different emotions on the basis of the speaker's facial expressions and the acoustic properties of speech (Fernald, 1993; Grossmann et al., 2006; Vaillant-Molina et al., 2013). Evidence using continuous speech typically shows that young infants rely on the congruence between auditory emotions (happy, angry) and the appropriate facial expressions (Soken and Pick, 1992; Walker-Andrews, 1997). Production-wise, young infants at 4–5 months of age express emotions such as sadness or enjoyment through facial expressions, and at 12 months of age their facial expressions can signal fear, pain, surprise, or interest (Sullivan and Lewis, 2003). At similar ages, vocal cues are also found to reflect their emotional states (Scheiner et al., 2002; Oller et al., 2013; Lindová et al., 2015).

It is not until much later, however, that children use this early sensitivity to visual and acoustic features of emotion to understand their interlocutor's affective state

(Nelson and Russell, 2011; Quam and Swingley, 2012; Berman et al., 2016). Berman et al. (2016) designed a task in which 3- and 5-year-old children had to match pictures of happy-looking and sad-looking faces to happy-sounding and sad-sounding speech, while explicit (pointing) and implicit (eye gaze) responses were measured. Results indicated that only 5 years old children were able to explicitly match the appropriate acoustic and visual cues of emotion, and that at 3 years of age they could only do it implicitly for the negative valence pair.

Even more challenging for children are stimuli in which the speaker intentionally mismatches the audiovisual cues of emotion from the contextual and lexical information, with the purpose of being ironic. In such cases, children at 5–6 years of age tend to interpret the utterance literally even if prosodic cues of emotion signal the speaker's irony (Nakassis and Snedeker, 2002; Laval and Bert-Erboul, 2005; Aguert et al., 2013; Bosco et al., 2013), and only if the utterance is produced together with visual cues of emotion can children infer non-literal meaning (Gil et al., 2014; González-Fuente, 2017). Taken together, all these findings indicate that vocal and visual cues of emotion are recognized and used very early in infancy, and that children use these early skills to process other people's emotions once more complex cognitive abilities are in place.

ACOUSTIC AND VISUAL MARKERS OF INTENTIONS, ATTITUDES, AND BELIEFS

Infants recognize and express their social intentions and communicative goals very early in development, and they use prosodic and gestural means to do so. Twelve-month-old infants rely on pitch, duration, and the shape of the gesture (open-palm pointing, index-finger pointing, etc.) to understand whether the interlocutor is communicating in order to request an object, to inform the caregiver about its presence, or to share interest about it (Behne et al., 2012; Sakkalou and Gattis, 2012; Esteve-Gibert et al., 2017c; Rohlfing et al., 2017). For example, 12-month-old infants use the shape of a pointing gesture and the information from the context to understand that their interlocutor is referring to a certain object in space with a specific social intention (Behne et al., 2012). Interestingly, when contextual cues are ambiguous or uninformative, 12-month-old infants use the shape of the pointing gesture in combination with the prosodic features of speech to infer the speakers' pragmatic intentions (Esteve-Gibert et al., 2017c). Some months later, at around 15 months of age, infants distinguish an action as being accidental or intentional only through the prosodic features of the interlocutor's speech (Sakkalou and Gattis, 2012).

At these pre-lexical stages of language development, prosody and gesture also enable infants to express their intentions toward their interlocutor. We know that 12-month-old infants produce pointing gestures toward referents in space with the purpose of requesting or declaring information, interest, attitudes, or actions (Tomasello et al., 2007; Kovács et al., 2014). It seems that not only pointing gestures but also the prosodic cues of the vocalizations accompanying them indicate the infants' intention (Grünloh and Liszkowski, 2015; Aureli et al., 2017). Aureli et al.

(2017), for instance, found that when Italian-learning 12- to 18-month-olds intend to produce points with a declarative function, the intonation of the vocalization accompanying these points is mostly falling, while it rises to accompany points aimed at asking objects from the interlocutor (thus paralleling what happens in adult speech).

The speaker's beliefs and attitudes about the content of the message are also signaled through vocal and visual strategies. Prosodic cues such as speech rate, pitch level and direction, or voice quality, and gestures such as eyebrow furrowing, head tilt, or shoulder shrugging, are reliably markers of the speaker being uncertain, incredulous, or polite (Krahmer and Swerts, 2005; Dijkstra et al., 2006; Crespo Sendra et al., 2013). Children need complex cognitive mental abilities (the so-called 'Theory of Mind') to understand and express these meanings in language (Wellman, 1990; Perner, 1991; Gopnik, 1993). A large body of research has dealt with the question of when these abilities emerge. Some researchers propose that only at ages four to five do children have fully developed mind-reading abilities, since it is at this age that they succeed in false-belief tasks (Wimmer and Perner, 1983; Baron-Cohen et al., 1985). Yet others claim that younger infants show early cognitive abilities of this kind when less cognitively demanding tasks are used (Onishi and Baillargeon, 2005; Baillargeon et al., 2010; Kovács et al., 2010). Studies exploring the development of prosodic and gesture cues to interpret the other's beliefs and attitudes suggest that children's belief comprehension increases significantly during the preschool years. For example, at 3–5 years of age children detect at above chance level the speaker's beliefs about what she/he is saying thanks to the speaker's facial expressions and, interestingly, those that are more accurate are those with more sophisticated belief-reasoning skills (Armstrong et al., 2014). Visual information is found to be a stronger cue for preschoolers than prosodic cues of uncertainty, even if prosody is a stronger indicator still than lexical information (Moore et al., 1993; Hübscher et al., 2017). On the production side, children first use prosody than lexical cues to mark uncertainty in speech (Hübscher et al., 2016), and at 7–8 year of age they signal uncertainty through facial expressions such as eyebrow raising or furrowing or funny faces, and with prosodic cues such as fillers, delays, and high intonation (Krahmer and Swerts, 2005; Visser et al., 2014). All together, these studies suggest that children use the acoustic and visual components of prosody before lexical markers to understand and produce beliefs and attitudes in language. Yet, more studies are required to disentangle which of these prosodic dimensions (visual or acoustic) comes first, and whether this developmental path depends on the child's cognitive abilities and/or on the specific linguistic meaning that is investigated.

DISCUSSION

The present review is aimed at highlighting recent discoveries on the developmental integration of speech in the auditory and visual domains, focusing on the prosodic level. Although there are more and more evidence of links between speech and gestures, we do not fully understand the relative weight of each

modality in language comprehension, and we need to clarify whether prosody has parallel forms and functions in the acoustic and visual domains. Adopting a developmental approach could help in answering these questions.

Developmental research can help disentangle whether gestures are part of the speakers' linguistic system. There is consistent evidence that infants and children first use the gesture modality to refer to objects in space before they use words and word-gesture combinations to do so (Bates et al., 1979; Butcher and Goldin-Meadow, 2000; Esteve-Gibert and Prieto, 2014). In fact, the rate of gesturally pointed referents is a reliable sign of the infants' vocabulary skills at later stages (Rowe and Goldin-Meadow, 2009; Igualada et al., 2015), and the rate of pointing-speech combinations at 18 months of age (when pointing and speech provide complementary meanings) is a reliable predictor of sentence complexity at 42 months of age (Rowe and Goldin-Meadow, 2009). Mathew et al. (2017) observed that 6-year-olds produce 'beat' gestures with an emphasizing function, but surprisingly the gesture-accompanying words did not always bear a pitch accent, suggesting that children are still learning to use the speech modality to emphasize discourse elements, while they seem to already master the gesture. Although not all language functions emerge first in the visual modality (note, for instance, that toddlers first express actions with verbs and only later are able to represent that same action with iconic gestures depicting that action; Özçaliskan et al., 2003), the abovementioned results indicate that infants and children do use gestures for linguistic purposes, and that speech and gestures might be part of the same linguistic and communicative system (Kendon, 1980; McNeill, 1992; Goldin-Meadow, 1998).

It is still an open question the reason why certain linguistic functions are first expressed through gestures and some others are first observed in the acoustic dimension. Paradé and Iverson (2011) propose a dynamic systems approach to cope with the fact that infants prefer to use one modality over the other for a given linguistic function at certain stages in language development. According to these authors, in periods where infants increase their skills in one communicative behavior, there might be a temporary regression in an alternative communicative behavior. For instance, the authors find that when infants' vocabulary increases, their production of multimodal communicative behaviors (i.e., combination of vocal, gestural, and affect behaviors) is reduced. Later, once vocabulary skills are stabilized, the rate of multimodal communicative behaviors increases again. It remains unclear, however, why certain linguistic functions emerge first through gesture rather than through speech, and vice-versa, as well as what motor, cognitive, or communicational factors might influence this behavior.

Studies in brain imagery could also help tease apart the possibility of a gesture/speech linkage in language. Indeed, in adult populations, it has been shown that listening to speech evokes neural responses in the motor cortex. This has been controversially interpreted as evidence that speech sounds are processed as articulatory movements (Pulvermüller and Fadiga, 2010). Recently, Biau et al. (2016) evaluated beat synchrony against arbitrary visual cues bearing equivalent

rhythmic and spatial properties as the gestures. Their results revealed that left Middle Temporal Gyrus and Inferior Frontal Gyrus were specifically sensitive to speech synchronized with beats, compared to the arbitrary vision–speech pairing. Hence, it seems that co-speech gestures and speech perception are instantiated through a specialized brain network sensitive to the communicative intent conveyed by the speaker's whole body.

There are very few studies investigating the developmental signs of the vocal-motor linkages at the neural level, and most evidence comes from populations with developmental disorders and brain injuries. For instance, children with perinatal brain lesions are found to have both lower rates of gesture production and smaller vocabularies (Sauer et al., 2010). Another way to specify the links between gestures and speech would be to explore how sensorimotor feedback influences auditory-visual speech processing, for instance by investigating whether the production of gestures influences infants' speech fluency. If more evidence is obtained showing that gesture and speech mutually influence each other in language production, perception, and comprehension, this would suggest that they are part of the linguistic system and not only communicative means, especially in development.

Among the linguistic aspects revealing the gesture/speech link more clearly, we have shown that prosody has a prominent status. Prosodic targets are anchoring points for manual gestures and facial expressions to align, pitch accents attracting prominent gestural phases and prosodic phrase boundaries framing the scope of gesture movements. This is true in adults (Hadar et al., 1983; De Ruiter, 1998; Leonard and Cummins, 2011; Esteve-Gibert and Prieto, 2013; Ferré, 2014; Ishi et al., 2014; Ambrazaitis and House, 2017; Esteve-Gibert et al., 2017a), and it also seems to hold for infants and children (Butcher and Goldin-Meadow, 2000; Esteve-Gibert and Prieto, 2014; Mathew et al., 2017). While more research is needed to examine the patterns of this temporal linkage in infants' productions (especially in stages when these prosodic targets become adult-like), perception studies show that infants are sensitive to the alignment of prosodic and visual cues as early as 8–9 months of age (Kitamura et al., 2014; Esteve-Gibert et al., 2015). It has been proposed that the driving force of this temporal linkage is a bi-directional influence between gesture and speech 'pulses' (i.e., peaks in an ongoing rhythm) (McNeill, 1992; Tuite, 1993; Iverson and Thelen, 1999; Port, 2003; Rusiewicz and Esteve-Gibert, 2018).

Prosody and gestures also overlap in terms of which linguistic functions they are used for. Infants use visual correlates of prosody to segment the speech stream (e.g., Kitamura et al., 2014;

Guellaï et al., 2016), to organize information at the discourse level (e.g., Nicoladis et al., 1999; Capone and McGregor, 2004; Mathew et al., 2017), and to express emotions, intentions, and beliefs (Sullivan and Lewis, 2003; Esteve-Gibert and Prieto, 2014; Berman et al., 2016; Aureli et al., 2017; González-Fuente, 2017). Children are sensitive to the fact that visual cues convey relevant linguistic meaning, and experimental evidence shows that gestures are processed earlier and more accurately than prosodic or lexical cues (Armstrong et al., 2014; Esteve-Gibert et al., 2017c; Hübscher et al., 2017). If future studies confirm that infants and children first process through visual cues what they later learn to process acoustically, this would mean that gestures are key in the development of linguistic categories, and that they not only precede but also scaffold language development (see a proposal on this regard in Hübscher et al., 2017). Furthermore, by examining in more detail how visual and acoustic cues of prosody emerge, evolve, and interact across development, we will be able to develop models that can predict and guide intervention in the case of atypical language development. The studies reviewed here have shown that gestures are tightly linked to prosody at the formal and functional levels and across different stages of language development. Still, further studies are needed to fully clarify the origin of these links and their implications for language acquisition.

AUTHOR CONTRIBUTIONS

All authors have equally participated to the discussion and writing of the manuscript. BG has had a leading role in section 1 (Introduction), section 3 (Word Learning), and section 7 (Discussion), while NE-G has had a leading role in section 2 (Temporal Aspects), section 4 (Narrative Skills), section 5 (Emotions), and section 6 (Intentions, Attitudes, and Beliefs).

FUNDING

This research was funded by the FJCI-2015-26845 postdoctoral grant (Spanish Ministry of Economy, Industry, and Competitiveness) to NE-G, and by the Fyssen Foundation for BG.

ACKNOWLEDGMENTS

We thank Pilar Prieto, Maya Gratier, and Alan Langus for their insights and discussion of the research presented in this article.

REFERENCES

- Aguert, M., Laval, V., Lacroix, A., Gil, S., and Le Bigot, L. (2013). Inferring emotions from speech prosody: not so easy at age five. *PLoS One* 8:e83657. doi: 10.1371/journal.pone.0083657
- Alamillo, A. R., Colletta, J. M., and Guidetti, M. (2013). Gesture and language in narratives and explanations: the effects of age and communicative activity on late multimodal discourse development. *J. Child Lang.* 40, 511–538. doi: 10.1017/S0305000912000062
- Aldridge, M. A., Braga, E. S., Walton, G. E., and Bower, T. G. R. (1999). The intermodal representation of speech in newborns. *Dev. Sci.* 2, 42–46. doi: 10.1111/1467-7687.00052
- Alibali, M. W., Kita, S., and Young, A. J. (2000). Gesture and the process of speech production: we think, therefore we gesture. *Lang. Cogn. Process.* 15, 593–613. doi: 10.1080/016909600750040571

- Ambrazaitis, G., and House, D. (2017). Multimodal prominences: exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Commun.* 95, 100–113. doi: 10.1016/j.specom.2017.08.008
- Armstrong, M., Esteve-Gibert, N., and Prieto, P. (2014). “The acquisition of multimodal cues to disbelief,” in *Proceedings of the 7th International Conference on Speech Prosody*, Dublin.
- Aureli, T., Spinelli, M., Fasolo, M., Garito, M. C., Perucchini, P., and D’Odorico, L. (2017). The pointing-vocal coupling progression in the first half of the second year of life. *Infancy* 22, 801–818. doi: 10.1111/inf.12181
- Baillargeon, R., Scott, R. M., and He, Z. (2010). False-belief understanding in infants. *Trends Cogn. Sci.* 14, 110–118. doi: 10.1016/j.tics.2009.12.006
- Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a ‘theory of mind’? *Cognition* 21, 37–46.
- Bates, E., Benigni, L., Bretherton, I., Camaioni, L., and Volterra, V. (1979). *The Emergence of Symbols: Cognition and Communication in Infancy*. New York, NY: Academic Press.
- Baumann, S., and Grice, M. (2006). The intonation of accessibility. *J. Pragmat.* 38, 1636–1657. doi: 10.1016/j.pragma.2005.03.017
- Behne, T., Liszkowski, U., Carpenter, M., and Tomasello, M. (2012). Twelve-month-olds’ comprehension and production of pointing. *Br. J. Dev. Psychol.* 30, 359–375. doi: 10.1111/j.2044-835X.2011.02043.x
- Berman, J. M. J., Chambers, C. G., and Graham, S. A. (2016). Preschoolers’ real-time coordination of vocal and facial emotional information. *J. Exp. Child Psychol.* 142, 391–399. doi: 10.1016/j.jecp.2015.09.014
- Berman, R. A., and Slobin, D. I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. Hillsdale, NJ: Erlbaum.
- Biau, E., Fernández, L. M., Holle, H., Avila, C., and Soto-Faraco, S. (2016). Hand gestures as visual prosody: BOLD responses to audio-visual alignment are modulated by the communicative nature of the stimuli. *Neuroimage* 132, 129–137. doi: 10.1016/j.neuroimage.2016.02.018
- Blossom, M., and Morgan, J. L. (2006). “Does the face say what the mouth says? A study of infants’ sensitivity to visual prosody,” in *Proceedings of the 30th Annual Boston University Conference on Language Development*, Somerville, MA.
- Booth, A. E., McGregor, K. K., and Rohlfing, K. L. (2008). Socio-pragmatics and attention: contributions to gesturally guided word learning in toddlers. *Lang. Learn. Dev.* 4, 179–202. doi: 10.1080/15475440802143091
- Bosco, F. M., Angeleri, R., Colle, L., Sacco, K., and Bara, B. G. (2013). Communicative abilities in children: an assessment through different phenomena and expressive means. *J. Child Lang.* 40, 741–778. doi: 10.1017/S0305000913000081
- Butcher, C., and Goldin-Meadow, S. (2000). “Gesture and the transition from one- to two-word speech: when hand and mouth come together,” in *Language and Gesture*, ed. D. McNeill (Chicago, IL: Cambridge University Press), 235–257.
- Byrd, D., and Saltzman, E. (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *J. Phon.* 31, 149–180. doi: 10.1016/S0095-4470(02)00085-2
- Capone, N. C., and McGregor, K. K. (2004). Gesture development: a review for clinical and research practices. *J. Speech Lang. Hear. Res.* 47, 173–186. doi: 10.1044/1092-4388(2004/015)
- Carlson, K., Clifton, C., and Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *J. Mem. Lang.* 45, 58–81. doi: 10.1006/jmla.2000.2762
- Chen, A. (2018). “Get the focus right across languages: acquisition of prosodic focus-marking in production,” in *The Development of Prosody in First Language Acquisition*, eds P. Prieto and N. Esteve-Gibert (Amsterdam: John Benjamins).
- Chong, S. C. F., Werker, J. F., Russell, J. A., and Carroll, J. M. (2003). Three facial expressions mothers direct to their infants. *Infant Child Dev.* 12, 211–232. doi: 10.1002/icd.286
- Chu, M., and Kita, S. (2011). The nature of gestures’ beneficial role in spatial problem solving. *J. Exp. Psychol. Gen.* 140, 102–115. doi: 10.1037/a0021790
- Clark, E. V., and Estigarribia, B. (2011). Using speech and gesture to introduce new objects to young children. *Gesture* 11, 1–23. doi: 10.1075/gest.11.1.01cla
- Colletta, J.-M., Guidetti, M., Capirci, O., Cristilli, C., Demir, O. E., Kunene-Nicolas, R. N., et al. (2014). Effects of age and language on co-speech gesture production: an investigation of French, American, and Italian children’s narratives. *J. Child Lang.* 42, 122–145. doi: 10.1017/S0305000913000585
- Cook, S. M., and Goldin-Meadow, S. (2006). The role of gesture in learning: Do children use their hands to change their minds? *J. Cogn. Dev.* 7, 211–232. doi: 10.1207/s15327647jcd0702_4
- Cooper, W. E., and Paccia-Cooper, J. (1980). *Syntax and Speech*. Cambridge, MA: Harvard University Press. doi: 10.4159/harvard.9780674283947
- Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.
- Crespo Sendra, V., Kaland, C., Swerts, M., and Prieto, P. (2013). Perceiving incredulity: the role of intonation and facial gestures. *J. Pragmat.* 47, 1–13. doi: 10.1016/j.pragma.2012.08.008
- Cristia, A. (2011). Fine-grained variation in caregivers’/s/ predicts their infants’/s/ category. *J. Acoust. Soc. Am.* 129, 3271–3280. doi: 10.1121/1.3562562
- Cutica, I., and Bucciarelli, M. (2008). The deep versus the shallow: effects of co-speech gestures in learning from discourse. *Cogn. Sci.* 32, 921–935. doi: 10.1080/03640210802222039
- Cvejic, E., Kim, J., and Davis, C. (2012). Recognizing prosody across modalities, face areas and speakers: examining perceivers’ sensitivity to variable realizations of visual prosody. *Cognition* 122, 442–453. doi: 10.1016/j.cognition.2011.11.013
- De Ruiter, J. P. (1998). *Gesture and Speech Production*. Doctoral dissertation, Katholieke Universiteit, Nijmegen.
- De Ruiter, J. P., Bangerter, A., and Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: investigating the tradeoff hypothesis. *Top. Cogn. Sci.* 4, 232–248. doi: 10.1111/j.1756-8765.2012.01183.x
- Diessel, H., and Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language* 81, 1–25. doi: 10.1353/lan.2005.0169
- Dijkstra, C., Krahmer, E., and Swerts, M. (2006). “Manipulating uncertainty: the contribution of different audiovisual prosodic cues to the perception of confidence,” in *Proceedings of the International Conference on Speech Prosody*, Dresden.
- Dohen, M., and Loevenbruck, H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Lang. Speech* 52, 177–206. doi: 10.1177/0023830909103166
- Ejiri, K. (1998). Relationship between rhythmic behavior and canonical babbling in infant vocal development. *Phonetica* 55, 226–237. doi: 10.1159/000028434
- Ejiri, K., and Masataka, N. (2001). Co-occurrence of preverbal vocal behavior and motor action in early infancy. *Dev. Sci.* 4, 6–11. doi: 10.1111/1467-7687.00147
- Ekman, P. (1979). “About brows: emotional and conversational signals,” in *Human Ethology: Claims and Limits of a New Discipline*, eds M. von Cranach, K. Foppa, W. Lepenies, and D. Ploog (Cambridge: Cambridge University Press), 169–202.
- Esteve-Gibert, N., Borràs-Comes, J., Asor, E., Swerts, M., and Prieto, P. (2017a). The timing of head movements: the role of prosodic heads and edges. *J. Acoust. Soc. Am.* 141, 4727–4739. doi: 10.1121/1.4986649
- Esteve-Gibert, N., Loevenbruck, H., Dohen, M., and D’Imperio, M. (2017b). “The use of prosody and gestures for the production of contrastive focus in French-speaking 4- and 5-year-old children,” in *Proceedings of the Workshop on Abstraction, Diversity and Speech Dynamics*, Munich.
- Esteve-Gibert, N., Liszkowski, U., and Prieto, P. (2016). “Prosodic and gestural features distinguish the intention of pointing gestures in child-directed communication,” in *Interdisciplinary Approaches to Intonational Grammar in Ibero-Romance*, eds M. E. Armstrong, N. Henriksen, and M. D. M. Vanrell (Amsterdam: John Benjamins), 251–275.
- Esteve-Gibert, N., and Prieto, P. (2013). Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *J. Speech Lang. Hear. Res.* 56, 850–864. doi: 10.1044/1092-4388(2012/12-0049)
- Esteve-Gibert, N., and Prieto, P. (2014). Infants temporally coordinate gesture-speech combinations before they produce their first words. *Speech Commun.* 57, 301–316. doi: 10.1016/j.specom.2013.06.006
- Esteve-Gibert, N., Prieto, P., and Liszkowski, U. (2017c). Twelve-month-olds understand social intentions based on prosody and gesture shape. *Infancy* 22, 108–129. doi: 10.1111/inf.12146
- Esteve-Gibert, N., Prieto, P., and Pons, F. (2015). Nine-month-old infants are sensitive to the temporal alignment of prosodic and gesture prominences. *Infant Behav. Dev.* 38, 126–129. doi: 10.1016/j.infbeh.2014.12.016

- Fernald, A. (1993). Approval and disapproval: infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Dev.* 64, 657–674. doi: 10.2307/1131209
- Fernald, A., and Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Dev. Psychol.* 20, 104–113. doi: 10.1037/0012-1649.20.1.104
- Ferré, G. (2014). A multimodal approach to markedness in spoken French. *Speech Commun.* 57, 268–282. doi: 10.1016/j.specom.2013.06.002
- Fisher, C., and Tokura, H. (1995). The given-new contract in speech to infants. *J. Mem. Lang.* 34, 287–310. doi: 10.1006/jmla.1995.1013
- Gil, S., Aguert, M., Le Bigot, L., Lacroix, A., and Laval, V. (2014). Children's understanding of others' emotional states. *Int. J. Behav. Dev.* 38, 539–549. doi: 10.1177/0165025414535123
- Gogate, L. J., Bahrick, L. E., and Watson, J. D. (2000). A study of multimodal motherese: the role of temporal synchrony between verbal labels and gestures. *Child Dev.* 71, 878–894. doi: 10.1111/1467-8624.00197
- Goldin-Meadow, S. (1998). *The Development of Gesture and Speech as an Integrated System*. San Francisco, CA: Jossey-Bass.
- González-Fuente, S. (2017). *Audiovisual Prosody and Verbal Irony*. Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona.
- Goodwyn, S. W., Acredolo, L. P., and Brown, C. A. (2000). Impact of symbolic gesturing on early language development. *J. Nonverbal Behav.* 24, 81–103. doi: 10.1023/A:1006653828895
- Gopnik, A. (1993). How we know our minds: the illusion of first-person knowledge of intentionality. *Behav. Brain Sci.* 16, 1–15. doi: 10.1017/S0140525X00028636
- Graziano, M. (2011). "The development of two pragmatic gestures of the so-called 'Open Hand Supine family' in Italian children," in *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon*, eds M. Seyfeddinipur and M. Gullberg (Amsterdam: John Benjamins Publishing Company), 311–330.
- Graziano, M. (2014). "Gestures in Southern Europe: children's pragmatic gestures in Italy," in *Body-Language-Communication*, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter), 1253–1258.
- Green, J. R., Nip, I. S., Wilson, E. M., Mefferd, A. S., and Yunusova, Y. (2010). Lip movement exaggerations during infant-directed speech. *J. Speech Lang. Hear. Res.* 53, 1529–1542. doi: 10.1044/1092-4388(2010/09-0005)
- Grieser, D. L., and Kuhl, P. K. (1988). Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Dev. Psychol.* 24, 14–20. doi: 10.1037/0012-1649.24.1.14
- Grossmann, T., Striano, T., and Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. *Dev. Sci.* 9, 309–315. doi: 10.1111/j.1467-7687.2006.00494.x
- Grünloh, T., and Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts. *J. Child Lang.* 42, 1312–1336. doi: 10.1017/S0305000914000816
- Guellai, B., Langus, A., and Nespor, M. (2014). Prosody in the hands of the speaker. *Front. Psychol.* 5:700. doi: 10.3389/fpsyg.2014.00700
- Guellai, B., Streri, A., Chopin, A., Rider, D., and Kitamura, C. (2016). Newborns' sensitivity to the visual aspects of infant-directed speech: evidence from point-line displays of talking faces. *J. Exp. Psychol. Hum. Percept. Perform.* 42, 1275–1281. doi: 10.1037/xhp0000208
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511616983
- Hadar, U., Steiner, T. J., Grant, E. C., and Rose, F. C. (1983). Kinematics of head movements accompanying speech during conversation. *Hum. Mov. Sci.* 2, 35–46. doi: 10.1016/0167-9457(83)90004-0
- Hübscher, I., Esteve-Gibert, N., Igualada, A., and Prieto, P. (2017). Intonation and gesture as bootstrapping devices in speaker uncertainty. *First Lang.* 37, 24–41.
- Hübscher, I., Vinzce, L., and Prieto, P. (2016). "Epistemic meaning is first communicated through gesture, face and prosody," in *Poster at the Workshop on Audiovisual Speech Processing and Language Learning*, Universitat Pompeu Fabra, Barcelona.
- Hudson, J., and Shapiro, L. (1991). "From knowing to telling: the development of children's scripts, stories, and personal narratives," in *Developing Narrative Structure*, eds A. McCabe and C. Peterson (Hillsdale, NJ: Lawrence Erlbaum Associates), 89–136.
- Igualada, A., Bosch, L., and Prieto, P. (2015). Language development at 18 months is related to communicative strategies at 12 months. *Infant Behav. Dev.* 39, 42–52. doi: 10.1016/j.infbeh.2015.02.004
- Igualada, A., Esteve-Gibert, N., and Prieto, P. (2017). Beat gestures improve word recall in 3- to 5-year-old children. *J. Exp. Child Psychol.* 156, 99–112. doi: 10.1016/j.jecp.2016.11.017
- Ishi, C. T., Ishiguro, H., and Hagita, N. (2014). Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Commun.* 57, 233–243. doi: 10.1016/j.specom.2013.06.008
- Iverson, J. M., and Fagan, M. K. (2004). Infant vocal-motor coordination: precursor to the gesture-speech system? *Child Dev.* 75, 1053–1066. doi: 10.1111/j.1467-8624.2004.00725.x
- Iverson, J. M., and Goldin-Meadow, S. (1998). Why do people gesture as they speak? *Nature* 396:228. doi: 10.1038/24300
- Iverson, J. M., and Thelen, E. (1999). Hand, mouth and brain. *J. Conscious. Stud.* 6, 19–40.
- Jesse, A., and Johnson, E. K. (2016). Audiovisual alignment of co-speech gestures to speech supports word learning in 2-year-olds. *J. Exp. Child Psychol.* 145, 1–10. doi: 10.1016/j.jecp.2015.12.002
- Jun, S.-A. (2005). "Prosodic typology," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. S.-A. Jun (Oxford: Oxford University Press), 430–458. doi: 10.1093/acprof:oso/9780199249633.003.0016
- Kallay, J., and Redford, M. A. (2016). "A longitudinal study of children's intonation in narrative speech," in *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, San Francisco, CA. doi: 10.21437/Interspeech.2016-1396
- Kelly, S., and Barr, D. (1999). Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. *J. Mem. Lang.* 40, 577–592. doi: 10.1006/jmla.1999.2634
- Kendon, A. (1980). "Gesticulation and speech: two aspects of the process of utterance," in *The Relationship of Verbal and Nonverbal Communication*, ed. M. R. Key (The Hague: Mouton De Gruyter), 207–227.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572
- Kim, J., Cvejic, E., and Davis, C. (2014). Tracking eyebrows and head gestures associated with spoken prosody. *Speech Commun.* 57, 317–330. doi: 10.1016/j.specom.2013.06.003
- Kitamura, C., and Burnham, D. (2003). Pitch and communicative intent in mother's speech: adjustments for age and sex in the first year. *Infancy* 4, 85–110. doi: 10.1207/S15327078IN0401_5
- Kitamura, C., Guellai, B., and Kim, J. (2014). Motherese by eye and ear: infants perceive visual prosody in point-line displays of talking heads. *PLoS One* 9:e111467. doi: 10.1371/journal.pone.0111467
- Kjelgaard, M. M., and Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *J. Mem. Lang.* 40, 153–194. doi: 10.1006/jmla.1998.2620
- Kovács, A. M., Tauzin, T., Teglas, E., Gergely, G., and Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy* 19, 543–557. doi: 10.1111/inf.12060
- Kovács, Á. M., Téglás, E., and Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science* 330, 1830–1834. doi: 10.1126/science.1190792
- Krahmer, E., and Swerts, M. (2005). How children and adults produce and perceive uncertainty in audiovisual speech. *Lang. Speech* 48, 29–53. doi: 10.1177/00238309050480010201
- Krahmer, E., and Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *J. Mem. Lang.* 57, 396–414. doi: 10.1016/j.jml.2007.06.005
- Krivokapic, J., Tiede, M. K., and Tyrone, M. E. (2017). A kinematic study of prosodic structure in articulatory and manual gestures: results from a novel method of data collection. *Lab. Phonol.* 8, 1–26. doi: 10.5334/labphon.75
- Krivokapic, J., Tiede, M. K., Tyrone, M. E., and Goldenberg, D. (2016). "Speech and manual gesture coordination in a pointing task," in *Proceedings of the 8th International Conference on Speech Prosody*, Boston, 1240–1244. doi: 10.21437/SpeechProsody.2016-255
- Kuhl, P. K., and Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behav. Dev.* 7, 361–381. doi: 10.1016/S0163-6383(84)80050-8
- Kuhl, P. K., Williams, K. A., and Meltzoff, A. N. (1991). Cross-modal speech perception in adults and infants using non-speech auditory stimuli. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 829–840. doi: 10.1037/0096-1523.17.3.829

- Ladd, R. (1996). *Intonational Phonology*, Cambridge Studies in Linguistics, Vol. 79. Cambridge: Cambridge University Press.
- Ladd, R. (2008). *Intonational Phonology*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511808814
- Langus, A., Marchetto, E., Bion, R. A., and Nespors, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *J. Mem. Lang.* 66, 285–306. doi: 10.1016/j.jml.2011.09.004
- Laval, V., and Bert-Erboul, A. (2005). French-speaking children's understanding of sarcasm: the role of intonation and context. *J. Speech Lang. Hear. Res.* 48, 610–620. doi: 10.1044/1092-4388(2005/042)
- Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa* 7, 102–122. doi: 10.1121/1.1982702
- Leonard, T., and Cummins, F. (2011). The temporal relation between beat gestures and speech. *Lang. Cogn. Process.* 26, 1457–1471. doi: 10.1080/01690965.2010.500218
- Levelt, W. J. M., Richardson, G., and La Heij, W. (1985). Pointing and voicing in deictic expressions. *J. Mem. Lang.* 24, 133–164. doi: 10.1016/0749-596X(85)90021-X
- Lindová, J., Špinková, M., and Nováková, L. (2015). Decoding of baby calls: can adult humans identify the eliciting situation from emotional vocalizations of preverbal infants? *PLoS One* 10:e0124317. doi: 10.1371/journal.pone.0124317
- Ma, W., Golinkoff, R. M., Houston, D. M., and Hirsh-Pasek, K. (2011). Word learning in infant and adult-directed speech. *Lang. Learn. Dev.* 7, 185–201. doi: 10.1080/15475441.2011.579839
- Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P., and Lee, C. S. (1992). Prosodic effects in minimal attachment. *Q. J. Exp. Psychol.* 45, 73–87. doi: 10.1080/14640794208401316
- Mathew, M., Yuen, I., and Demuth, K. (2017). Talking to the beat: six-year-olds' use of stroke-defined non-referential gestures. *First Lang.* (in press). doi: 10.1177/0142723717734949
- McGregor, K. K., Rohlfing, K., Bean, A., and Marschner, E. (2009). Gesture as a support for word learning: the case of under. *J. Child Lang.* 36, 807–828. doi: 10.1017/S0305000908009173
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago.
- McNeill, D. (2005). *Gesture and Thought*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226514642.001.0001
- Moore, C., Harris, L., and Patriquin, M. (1993). Lexical and prosodic cues in the comprehension of relative certainty. *J. Child Lang.* 20, 153–167. doi: 10.1017/S030500090000917X
- Nagel, H. N., Shapiro, L. P., Tuller, B., and Nawy, R. (1996). Prosodic influences on the resolution of temporary ambiguity during on-line sentence processing. *J. Psycholinguist. Res.* 25, 319–344. doi: 10.1007/BF01708576
- Nakassis, C., and Snedeker, J. (2002). "Beyond sarcasm: intonation and context as relational cues in children's recognition of irony," in *Proceedings of the Annual Boston-University Conference on Language Development*, Vol. 26, eds A. Greenhill, M. Hughs, H. Littlefield, and H. Walsh (Somerville, MA: Cascadia Press), 429–440.
- Nelson, N. L., and Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *J. Exp. Child Psychol.* 110, 52–61. doi: 10.1016/j.jecp.2011.03.014
- Nespor, M., and Sandler, W. (1999). Prosody in Israeli sign language. *Lang. Speech* 42, 143–176. doi: 10.1177/00238309990420020201
- Nespor, M., and Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris, 327.
- Nespor, M., and Vogel, I. (2007). *Prosodic Phonology*, 1st Edn. Berlin: Mouton De Gruyter. doi: 10.1515/978311097790
- Nicoladis, E., Mayberry, R. L., and Genesee, F. (1999). Gesture and early bilingual development. *Dev. Psychol.* 35, 514–526. doi: 10.1037/0012-1649.35.2.514
- Oller, D. K. (2000). *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum.
- Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., and Chorna, L. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6318–6323. doi: 10.1073/pnas.1300337110
- O'Neill, M., Bard, K. A., Linnell, M., and Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Dev. Sci.* 8, 352–359. doi: 10.1111/j.1467-7687.2005.00423.x
- Onishi, K. H., and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* 308, 255–258.
- Özcaliskan, S., Gentner, D., and Goldin-meadow, S. (2013). Do iconic gestures pave the way for children's early verbs? *Appl. Psycholinguist.* 35, 1143–1162.
- Parladé, M. V., and Iverson, J. M. (2011). The interplay between language, gesture, and affect during communicative transition: a dynamic systems approach. *Dev. Psychol.* 47, 820–833. doi: 10.1037/a0021811
- Patterson, M. L., and Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behav. Dev.* 22, 237–247. doi: 10.1016/j.cognition.2008.05.009
- Patterson, M. L., and Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *J. Exp. Child Psychol.* 81, 93–115. doi: 10.1006/jecp.2001.2644
- Patterson, M. L., and Werker, J. F. (2003). Two-month-old infants match phonemic information in lips and voice. *Dev. Sci.* 6, 191–196. doi: 10.1016/j.cognition.2008.05.009
- Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.
- Pierrehumbert, J., and Hirschberg, J. (1990). "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, eds P. R. Cohen, J. Morgan, and M. E. Pollack (Cambridge, MA: MIT Press), 270–311.
- Ping, R., and Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cogn. Sci.* 34, 602–619. doi: 10.1111/j.1551-6709.2010.01102.x
- Poggi, I., D'Errico, F., and Vincze, L. (2013). Comments by words, face and body. *J. Multimodal User Interfaces* 7, 67–78. doi: 10.1007/s12193-012-0102-z
- Port, R. F. (2003). Meter and speech. *J. Phon.* 31, 599–611. doi: 10.1016/j.wocn.2003.08.001
- Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. *J. Acoust. Soc. Am.* 90, 2956–2970. doi: 10.1121/1.401770
- Prieto, P., Pugliesi, C., Borràs-Comes, J., Arroyo, E., and Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *J. Phon.* 49, 41–54. doi: 10.1016/j.wocn.2014.10.005
- Puccini, D., and Liszkowski, U. (2012). 15-month-old infants fast map words but not representational gestures of multimodal labels. *Front. Psychol.* 3:101. doi: 10.3389/fpsyg.2012.00101
- Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360. doi: 10.1038/nrn2811
- Quam, C., and Swingle, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Dev.* 83, 236–250. doi: 10.1111/j.1467-8624.2011.01700.x
- Rauscher, F. H., Krauss, R. M., and Chen, Y. (1996). Gesture, speech and lexical access: the role of lexical movements in speech production. *Psychol. Sci.* 7, 226–231. doi: 10.1002/wcs.1211
- Rohlfing, K. J., Grimminger, A., and Lüke, C. (2017). An interactive view on the development of deictic pointing in infancy. *Front. Psychol.* 8:1319. doi: 10.3389/fpsyg.2017.01319
- Roustan, B., and Dohen, M. (2010). "Co-production of contrastive focus and manual gestures: temporal coordination and effects on the acoustic and articulatory correlates of focus," in *Proceedings of the International Conference on Speech Prosody*, Chicago, IL.
- Rowe, M. L., and Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Dev. Sci.* 12, 182–187. doi: 10.1111/j.1467-7687.2008.00764.x
- Rusiewicz, H. L., and Esteve-Gibert, N. (2018). "Set in time: temporal coordination of prosody and gesture in the development of spoken language production," in *The Development of Prosody in First Language Acquisition*, eds P. Prieto and N. Esteve-Gibert (Amsterdam: John Benjamins).
- Sakkalou, E., and Gattis, M. (2012). Infants infer intentions from prosody. *Cogn. Dev.* 27, 1–16. doi: 10.1016/j.cogdev.2011.08.003
- Sandler, W. (1999). Prosody in two natural language modalities. *Lang. Speech* 42, 127–142. doi: 10.1177/00238309990420020101
- Sandler, W. (2005). Prosodic constituency and intonation in sign language. *Linguist. Ber.* 13, 59–86.

- Sandler, W. (2011). "The phonology of movement in sign language," in *Blackwell Companion to Phonology*, eds M. van Oostendorp, C. Ewen, K. Rice, and E. Hume (Oxford: Wiley-Blackwell), 557–603.
- Sandler, W. (2012). Dedicated gestures the emergence of sign language. *Gesture* 12, 265–307. doi: 10.1075/gest.12.3.01san
- Sauer, E., Levine, S. C., and Goldin-Meadow, S. (2010). Early gesture predicts language delay in children with pre- or perinatal brain lesions. *Child Dev.* 81, 528–539. doi: 10.1111/j.1467-8624.2009.01413.x
- Scheiner, E., Hammerschmidt, K., Jürgens, U., and Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *J. Voice* 16, 509–529. doi: 10.1016/S0892-1997(02)00127-3
- Soken, N. H., and Pick, A. D. (1992). Intermodal perception of happy and angry expressive behaviors by seven-month-old infants. *Child Dev.* 63, 787–795. doi: 10.2307/1131233
- Song, J. Y., Demuth, K., and Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *J. Acoust. Soc. Am.* 128, 389–400. doi: 10.1121/1.3419786
- Sullivan, M. W., and Lewis, M. (2003). Contextual determinants of anger and other negative expressions in young infants. *Dev. Psychol.* 39, 693–705. doi: 10.1037/0012-1649.39.4.693
- Sumby, W. H., and Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26, 212–215. doi: 10.1121/1.1907309
- Swerts, M., and Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Lang. Speech* 37, 21–43. doi: 10.1177/002383099403700102
- Swerts, M., and Krahmer, E. (2010). Visual prosody of newsreaders: effects of information structure, emotional content and intended audience on facial expressions. *J. Phon.* 38, 197–206. doi: 10.1016/j.wocn.2009.10.002
- Tomasello, M., Carpenter, M., and Liszkowski, U. (2007). A new look at infant pointing. *Child Dev.* 78, 705–722. doi: 10.1111/j.1467-8624.2007.01025.x
- Tuite, K. (1993). The production of gesture. *Semiotica* 93, 83–105. doi: 10.1515/semi.1993.93.1-2.83
- Vaillant-Molina, M., Bahrick, L. E., and Flom, R. (2013). Young infants match facial and vocal emotional expressions of other infants. *Infancy* 18, 1–15. doi: 10.1111/inf.12017
- Vihman, M. M., DePaolis, R. A., and Keren-portnoy, T. (2009). "A dynamic systems approach to babbling and words," in *The Cambridge Handbook of Child Language*, ed. E. L. Bavin (Cambridge: Cambridge University Press), 163–182.
- Vilà-Giménez, I., Igualada, A., and Prieto, P. (2017). *The Positive Effect of Observing and Producing Beat Gestures on Children's Narrative Abilities*. *Architectures and Mechanisms of Language Processing (AMLaP)*. Lancaster: Lancaster University.
- Visser, M., Krahmer, E., and Swerts, M. (2014). Children's expression of uncertainty in collaborative and competitive contexts. *Lang. Speech* 57, 86–107. doi: 10.1177/0023830913479117
- Wagner, P., Malisz, Z., and Kopp, S. (2014). Gesture and speech in interaction: an overview. *Speech Commun.* 57, 209–232. doi: 10.1016/j.specom.2013.09.008
- Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviours: differentiation of multimodal information. *Psychol. Bull.* 121, 437–456. doi: 10.1037/0033-2909.121.3.437
- Warren, P., Grabe, E., and Nolan, F. (1995). Prosody, phonology and parsing in closure ambiguities. *Lang. Cogn. Process.* 10, 457–486. doi: 10.1080/01690969508407112
- Weber, A., Braun, B., and Crocker, M. W. (2006). Finding referents in time: eye-tracking evidence for the role of contrastive accents. *Lang. Speech* 49, 367–392. doi: 10.1177/00238309060490030301
- Wellman, H. M. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.
- Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition* 103, 147–162. doi: 10.1016/j.cognition.2006.03.006
- Wimmer, H., and Perner, J. (1983). Belief about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128. doi: 10.1016/0010-0277(83)90004-5
- Yasinnik, Y., Renwick, M., and Shattuck-Hufnagel, S. (2004). "The timing of speech-accompanied gestures with respect to prosody," in *Proceedings of the From Sound to Sense Conference* (Cambridge, MA: MIT), 97–102.
- Zammit, M., and Schafer, G. (2011). Maternal label and gesture use affects acquisition of specific object names. *J. Child Lang.* 38, 201–221. doi: 10.1017/S0305000909990328

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Esteve-Gibert and Guellai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Production and Comprehension of Prosodic Markers in Sign Language Imperatives

Diane Brentari^{1*}, Joshua Falk¹, Anastasia Giannakidou¹, Annika Herrmann², Elisabeth Volk³ and Markus Steinbach³

¹ Department of Linguistics, Center for the Study of Gesture, Sign and Language, University of Chicago, Chicago, IL, United States, ² Department Language, Literature, Media I, Institute for German Sign Language, University of Hamburg, Hamburg, Germany, ³ Sign Language Lab and RTG 2070 "Understanding Social Relationships", Department of German Philology, University of Göttingen, Göttingen, Germany

OPEN ACCESS

Edited by:

Wendy Sandler,
University of Haifa, Israel

Reviewed by:

Ronnie Bring Wilbur,
Purdue University, United States

Onno Crasborn,
Radboud University Nijmegen,
Netherlands

Dachkovsky Svetlana,
University of Haifa, Israel

*Correspondence:

Diane Brentari
dbrentari@uchicago.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 16 November 2017

Accepted: 30 April 2018

Published: 23 May 2018

Citation:

Brentari D, Falk J, Giannakidou A, Herrmann A, Volk E and Steinbach M (2018) Production and Comprehension of Prosodic Markers in Sign Language Imperatives. *Front. Psychol.* 9:770. doi: 10.3389/fpsyg.2018.00770

In signed and spoken language sentences, imperative mood and the corresponding speech acts such as for instance, command, permission or advice, can be distinguished by morphosyntactic structures, but also solely by prosodic cues, which are the focus of this paper. These cues can express paralinguistic mental states or grammatical meaning, and we show that in American Sign Language (ASL), they also exhibit the function, scope, and alignment of prosodic, linguistic elements of sign languages. The production and comprehension of prosodic facial expressions and temporal patterns therefore can shed light on how cues are grammaticalized in sign languages. They can also be informative about the formal semantic and pragmatic properties of imperative types not only in ASL, but also more broadly. This paper includes three studies: one of production (Study 1) and two of comprehension (Studies 2 and 3). In Study 1, six prosodic cues are analyzed in production: temporal cues of sign and hold duration, and non-manual cues including tilts of the head, head nods, widening of the eyes, and presence of mouthings. Results of Study 1 show that neutral sentences and commands are well distinguished from each other and from other imperative speech acts via these prosodic cues alone; there is more limited differentiation among explanation, permission, and advice. The comprehension of these five speech acts is investigated in Deaf ASL signers in Study 2, and in three additional groups in Study 3: Deaf signers of German Sign Language (DGS), hearing non-signers from the United States, and hearing non-signers from Germany. Results of Studies 2 and 3 show that the ASL group performs significantly better than the other 3 groups and that all groups perform above chance for all meaning types in comprehension. Language-specific knowledge, therefore, has a significant effect on identifying imperatives based on targeted cues. Command has the most cues associated with it and is the most accurately identified imperative type across groups—indicating, we suggest, its special status as the strongest imperative in terms of addressing the speaker's goals. Our findings support the view that the cues are accessible in their content across groups, but that their language-particular combinatorial possibilities and distribution within sentences provide an advantage to ASL signers in comprehension and attest to their prosodic status.

Keywords: imperatives, speech acts, sign languages, gesture, prosody, semantics, non-manual markers

INTRODUCTION

It is well known that signers use their hands, body, head, and face for both grammatical and gestural purposes (see Goldin-Meadow and Brentari, 2017 for a review), and non-manual markers have been identified in signed and spoken languages to express sentence meaning, as well as emotion, intention, and the mental states of signers and speakers (for sign languages see Baker-Shenk, 1983; Poizner et al., 1987; Emmorey, 1999; Wilbur, 2003; Dachkovsky and Sandler, 2009; Dachkovsky et al., 2013; for good summaries see Pfau and Quer, 2010; Sandler, 2012; for spoken languages see Bolinger, 1983; Borràs-Comes and Prieto, 2011; Borràs-Comes et al., 2014; Domaneschi et al., 2017). In this paper, presenting three studies, we analyze the temporal and non-manual prosodic cues associated with imperative constructions as expressed in American Sign Language (ASL). We argue that, while the non-manuals may be comprehensible to non-signers to a large extent, in a sign language they take on specific distributions as part of sign language prosody and achieve grammatical status (cf. Nespor and Sandler, 1999; Wilbur, 1999, 2009, 2011, 2018; Sandler and Lillo-Martin, 2006; Brentari et al., 2011, 2015; Sandler, 2012).

Our motivation for this group of studies is 2-fold. First, we want to better interpret how signers and non-signers understand the prosodic cues of a sign language in the expression of speech acts, especially non-manuals that may also be comprehensible by non-signers to some extent (for prosodic cues of speech acts in spoken languages see Hellbernd and Sammler, 2016). Because of their pragmatic status as directive speech acts, imperatives engage a number of expressive facial expressions that have this intermediate status. To address this question we carefully annotate and analyze the cues produced in ASL imperatives in a production study (Study 1), and we then use those productions as stimuli for two comprehension studies (Studies 2 and 3), which include two groups of Deaf signers—ASL and German Sign Language (DGS)—as well as two hearing groups of non-signers from the United States and Germany. If all groups are able to perform equally well on a task of imperative comprehension, this would lead us to conclude that the prosodic cues and their patterns in ASL are equally accessible to signers and non-signers, while any differences among the groups would allow us to infer that modality- and language-specific experience, or effects of a specific gestural competence within a community, can affect accessibility. A second motivation for these studies is to shed light on the semantics that underlie imperative types (i.e., the imperative sentence mood). We investigate which imperative speech acts are most clearly distinguished in production, and which are most easily comprehensible across groups.

Sign Language Prosody

Prosody in sign languages takes the form of temporal properties of a word or phrase and accompanying non-manual cues. As in spoken languages, sign language prosody is relevant at several grammatical levels. At the lexical level, Wilbur (1999) has argued that the telicity of verbs and phrasal prominence are expressed by the prosodic properties of acceleration and deceleration, and this type of prominence can be paired with a number of different

sign types. The temporal markers of sign language prosody, such as lengthening a sign's duration or final hold together with non-manual edge markers such as head nods and eye blinks have been argued to mark constituent boundaries (cf. Nespor and Sandler, 1999; Sandler, 1999; Brentari et al., 2011, 2015).

Linguistic and gestural descriptions of non-manual markers refer to aspects such as position, movement, tension, aperture, and duration of musculature of the face, the head, and the body. Grammatical non-manual markers combine simultaneously with manual components as well as with other non-manual markers; that is, the grammatical information marked by non-manuals can be layered in a complex fashion (cf. Wilbur, 2000; Dachkovsky and Sandler, 2009). Non-manuals are also capable of deriving compositional meaning (cf. Nespor and Sandler, 1999; Sandler and Lillo-Martin, 2006; Herrmann, 2015). Sandler and Lillo-Martin (2006) and Dachkovsky and Sandler (2009) demonstrated that factual conditional sentences in Israeli Sign Language (ISL) are marked by brow raise, whereas counterfactual conditionals require an additional squint. They argue that each non-manual marker has inherent semantic properties, which are compositionally combined to derive the more complex counterfactual meaning. In other cases, the layering of non-manual markers is ascribed to the strong physical relation of the specific components, which jointly fulfill the same grammatical function. This applies, for example, to the forward head tilt in polar questions of DGS, which is regularly accompanied by a forward body lean (cf. Herrmann and Pendzich, 2014). A further distinction is commonly drawn between the upper face and the lower face among grammatical non-manuals (cf. Liddell, 1980; Coerts, 1992; Wilbur, 2000) whereby the upper face includes movements of the eyes and brows, and has been grammatically associated with larger units of prosody (phrases, clauses, utterances) whereas movements and positions of the lower face and mouth have been associated with smaller prosodic units, such as the syllable and prosodic word (Wilbur, 2000; Brentari and Crossley, 2002). In the studies we present here, the focus is on the eyes, head, and the presence or absence of "mouthings," which are the partial, silent articulations of an English word.

With regard to the grammatical structure of sign languages, non-manual markers play an essential role at all levels of grammar. Starting with phonology, non-manual markers can be lexically specified representing an inherent phonological feature of an individual sign (cf. Coerts, 1992; Brentari, 1998; Woll, 2001; Liddell, 2003; Pfau and Quer, 2010; Pendzich, 2016). The sign *RECENTLY* in DGS is, for instance, produced with a slightly protruded tip of the tongue (cf. Herrmann and Pendzich, 2014). The sign *RECENTLY* in ASL requires a subtle sideward head turn and tensed cheek muscles (cf. Liddell, 1980). Moreover, non-manual markers operate on the level of morphology expressing adverbial and adjectival meanings (cf. Liddell, 1980; Vogt-Svendsen, 2001; Pfau and Quer, 2010). A specific non-manual configuration in ASL using a tongue-thrust functions as an adverbial modifier meaning *carelessly* (cf. Liddell, 1980, p. 50).

Non-manual markers can also affect phrasal and sentence meaning and spread over larger prosodic domains such as phonological and intonational phrases (cf. Sandler, 2010;

Crasborn and van der Kooij, 2013; Herrmann, 2015; among others). The prosodic component is autonomous in the grammar and has been shown to interface with the semantics and pragmatics of sign languages (Sandler and Lillo-Martin, 2006; Sandler, 2010). While non-manuals can indicate syntactic constituency when combined with certain types of signs, as in the case of relative clauses in ASL (Liddell, 1980), the timing and spreading behavior of non-manual markers is associated with prosodic constituency, in particular with the intonational phrase (cf. Nespor and Sandler, 1999; Sandler, 2010). As demonstrated in (1) from ISL (Meir and Sandler, 2008, p. 153), the intonational phrase is not necessarily isomorphic with the syntactic domain. The whole sentence represents a polar question syntactically, but the non-manual marker *brow raise* (br) is argued to correspond to a rising question intonation, which only scopes over the first conjunct (cf. Sandler, 2010).

- (1) The prosodic, rather than syntactic domains of prosodic cues in ISL (Meir and Sandler, 2008)

br

YOU WANT ICE CREAM WHITE IX-A OR CHOCOLATE IX-B
[ISL]
'Do you want vanilla ice cream or chocolate?'

In addition to grammatical non-manuals, such as those described above, both spoken and signed communication also involve facial expressions to express affective meanings and mental states (cf. Campbell et al., 1999; Keltner et al., 2003; McCullough et al., 2005), which we will refer to as "expressives." Several affective facial expressions associated with a set of basic emotional states such as anger, sadness, or joy are claimed to be universal and therefore cross-culturally conveyed in a similar way (cf. Ekman and Friesen, 1971; Izard, 1994; Benitez-Quiroz et al., 2016). Affective facial expressions include several types, however. One type involves evaluative meaning as the expression of mental states, such as "surprise" or "puzzlement" (cf. Campbell, 1997; Emmorey, 1999).

Another important group of gestural non-manual markers are "iconic" and mimetic mouth gestures (cf. Sandler, 2009). Some iconic mouth gestures are comparable to manual iconic co-speech gestures (cf. McNeill, 1992), since they are produced simultaneously with signs and convey properties of objects or events. Accordingly, Sandler (2009) demonstrates that iconic mouth gestures can be used in narration settings to embellish or complement the linguistic descriptions produced by the hands. In her study of ISL, signers used a manual classifier construction to depict the journey of a cat up a drainpipe, while one of several ways to indicate the narrowness of the pipe was a tightened mouth movement, and one of several ways to indicate a bend in the drainpipe was a zigzag mouth movement, which aligned with the manual linguistic descriptions. These iconic forms were variable across signers and this was one of the reasons for considering them gestural. We consider expressive and iconic forms to be different types of non-manuals. Both may be accessible to non-signers to some degree, but the former refers to the speaker's or signer's affect (or quoted speaker's or signer's affect). Iconic forms refer to

the properties of an entity (e.g., size, shape) or an event (manner).

Even though grammatical and affective non-manuals share the same articulatory bases, they are argued to differ in a number of important ways. Experimental evidence has been helpful in distinguishing grammatical and affective non-manual markers. For instance, in grammaticality judgment tasks, signers have clear intuitions about grammatical non-manuals. By contrast, affective non-manuals result in greater within-individual variability (cf. Baker-Shenk, 1983; Poizner et al., 1987; Emmorey, 1999; Wilbur, 2003). In addition, McCullough and Emmorey (2009) investigated whether stimuli of continuously varying facial expressions are perceived categorically, i.e., whether they result in categorical perception (CP) effects. They found that sign language experience influences CP effects for grammatical, but not affective, non-manuals. Further evidence for distinctive representations of grammatical and affective non-manuals is based on neuropsychological studies, which demonstrate that grammatical facial expressions are processed in the left hemisphere, whereas affective facial expressions activate areas in the right hemisphere of the brain (cf. Poizner et al., 1987; Corina et al., 1999; McCullough et al., 2005; Corina and Spotswood, 2012). Finally, research on sign language acquisition reveals that Deaf infants are competent in using a set of affective non-manual markers such as the side-to-side headshake or brow furrow in both production and perception at an early age, but their grammatical use appears later during acquisition. Moreover, when both a manual and non-manual marker have the same grammatical function, such as in conditionals in ASL, the manual marker is acquired first (cf. Baker-Shenk, 1983; Reilly et al., 1990; Emmorey et al., 1995; Morgan and Woll, 2002; Reilly and Anderson, 2002; Brentari et al., 2015).

Grammatical and affective non-manuals differ in their distribution in terms of on- and offset, scope, and apex (cf. Liddell, 1980; Corina et al., 1999; Wilbur, 2003; Dachkovsky, 2007). Accordingly, the on- and off-set of grammatical non-manual markers align with consistent phrasal boundaries, and grammatical non-manuals display a sudden increase of intensity and have an abrupt onset and offset. In other words, the scope and source of grammatical non-manuals are linguistically defined. In contrast, non-manuals expressing emotional and evaluative states that do not contribute to the linguistic meaning display a gradual on- and offset as well as more variable spreading behavior (cf. de Vos et al., 2009), and the apex of intensity also allows for more variability (cf. Liddell, 1980).

One widely held view about non-manual marking is that emotional and mental states, iconic depictions, and discursive functions may be more accessible to non-signers, while grammatical markers may be more arbitrary and inaccessible (cf. Herrmann and Pendzich, 2014). However, even grammatical facial expressions in sign languages have varying degrees of accessibility, as seen in examples (2) and (3)—ranging from those largely accessible to non-signers [e.g., head nod to mean positive assertion (2a) vs. headshake to mean negation (2b)] to those that are relatively inaccessible (e.g., conditionals; see also Malaia and Wilbur, 2012; Malaia et al., 2013; Strickland et al., 2015). Even if negative headshake has language-particular distributional

properties that are relevant for the syntactic, typological groups of sign languages (cf. Quer, 2012), both head nods for assertion and headshake for negation are also accessible to non-signers. In contrast, the difference between two simple conjoined clauses in (3a) and a complex conditional construction in (3b) is thought to be less accessible to non-signers. Both clauses have neutral expressions, but in (3a), the neutral expression is extended over both clauses, while in (3b), the first clause has a brow raise used for conditionals. These four sentences are included in the Supplementary Materials.

(2) Sentence meanings that are relatively accessible to non-signers

a. Assertion (**Video 1**, Supplementary Materials)

head nod

I GO

"I'm going."

b. Negation (**Video 2**, Supplementary Materials)

headshake

I GO

"I'm not going."

(3) Sentence meanings that are relatively inaccessible to non-signers

a. Coordinate Structure (**Video 3**, Supplementary Materials)

neutral

YOU COME I LEAVE

"You come and I'll leave."

b. Conditional Structure (**Video 4**, Supplementary Materials)

brow raise neutral

YOU COME I LEAVE

"If you come, I'll leave."

This investigation targets the moment at which affective/expressive forms take on systematic linguistic distributions. As expressives, they may only be *paralinguistic* (Bolinger, 1983), and if that is the case there should be no advantage for knowing the grammar of ASL; however, if they have a systematic function, scope, and alignment in production and are used to the advantage of the ASL signers, we have evidence for their systematically linguistic status as part of the prosodic system. We will argue that the temporal and non-manual properties of the expressives associated with imperatives that we investigate in this paper are grammatical and prosodic, even though they may be accessible to non-signers, since they scope over and align with specific phrases, add prominence and also suprasegmental meaning, and have a semantic and syntactic role as well.

Imperatives

We now turn to the semantic and pragmatic properties of imperatives. Following recent analyses of imperatives (Portner, 2007; Condoravdi and Lauer, 2012; Kaufmann, 2012; von Fintel and Iatridou, 2017), we assume that imperative sentence types are associated with a conventionalized meaning, the "imperative mood." The imperative meaning appears to be flexible, and is

compatible with a range of speech acts such as for example, command, warning, and permission. Across languages, both spoken and signed, imperatives employ prosodic cues along with lexical and morphological markers, such as particles, word order or verbal inflection, and imperatives can also be expressed by prosodic cues alone (see Iatridou, 2008; Hellbernd and Sammler, 2016; von Fintel and Iatridou, 2017 for spoken languages; Donati et al., 2017 for sign languages). Donati et al. (2017) present an in-depth study of imperatives in three sign languages—Italian Sign Language (LIS), French Sign Language (LSF), and Catalan Sign Language (LSC). They found that a number of manual signs, as well as temporal and non-manual markers were used to express different types of imperatives cross-linguistically. While there are more than four pragmatic types of imperatives studied in Donati et al. (2017), in the current studies we focus on the four speech acts expressed by the imperatives described in (4). As we discuss below, these four speech acts belong to the group of illocutionary forces typically realized with imperatives. At the same time, the contextual conditions on these four speech acts are different enough to clearly distinguish these speech acts from each other¹.

(4) Imperatives and example contexts

a. Commands: You must do 'x'.

Possible context: You and a friend are in a library and you are trying to hurry your friend along. You say, "*Find a book*, and let's go."

b. Explanation: You must do 'x' in order to achieve some goal.

Possible context: You and a friend are in a library and you are explaining how to borrow a book. You say, "*Find a book*, take it to the desk, show your card, and allow them to stamp the book with the due date."

c. Permission: You may do 'x'.

Possible context: You and a friend are in a library and you agree to allow her to borrow a book with your library card, since she does not have one. You say, "*Find a book* and use my card."

d. Advice: You ought to or may do 'x' if you want to achieve some goal.

Possible context: You and a friend are in a library and your friend asks for advice on how to fix her car. You explain that you don't have that type of expertise, but since she is in a library you say, "*Find a book* and figure it out on your own."

The examples in (4)—which can be enriched with distinctions such as requests, exhortations, prohibitions, etc., and which can be understood as falling in one of the subtypes identified—illustrate that imperatives can be used to realize quite different speech acts. There is a lot of discussion about the types of speech acts expressed by imperatives in the semantics and

¹In addition, we wanted to study the temporal and non-manual markers of imperatives, and thus the sentences (i.e., the signs themselves and their order) had to be the same across all speech acts. As this is not the case for all types of imperatives, we avoided imperatives that involve different signs or different sign orders.

pragmatics literature (see Portner, 2005; Kaufmann, 2012 for recent overviews), and often the question is raised whether the imperative has a uniform meaning. The obvious variation illustrated above suggests that the imperative is flexible in illocutionary force, but it is also generally recognized that the command is the “prototypical” use of the imperative (as the word itself suggests). Considering these four imperative types, the command is relatively more important from the perspective of the speaker since it is the only speech act of the four driven primarily by the speaker’s goal or needs. The speaker has authority and uses the imperative as a command to get the addressee to do something the speaker wants or deems necessary. The other three imperative types (explanation, permission, and advice) take the perspective of the addressee, and involve primarily addressee goals, i.e., the imperative is used to further a goal of the addressee. By their very nature, then, addressee-goal imperatives appear weaker. This difference has not been featured prominently in the literature, but it is instrumental, as it turns out, to understanding the distinctive pattern of command we observe in our studies.

The different types of imperative speech acts are derived at the semantics/pragmatics interface on the basis of different contextual conditions² Framing our observations in terms of preference (Condoravdi and Lauer, 2012), in commands such as (4a), the speaker has a very strong preference for the addressee to find a book, and the addressee knows that he/she is responsible for the realization of the preference. Something similar holds for explanations, although it is in the interest of the addressee (rather than the speaker) to follow through. In permissions such as (4c), it is not the speaker but the addressee who has a preference to find a book. In this context, the imperative expresses a change of the speaker preference to the preference of the addressee. Likewise, in speech acts of advice such as (4d), the speaker either has a weak preference for the addressee to find a book or the speaker may add the preference of the addressee to find a book to his/her preference. Either way, the speaker does not have a strong preference for the proposition expressed by an imperative of advice and addresses only the goal of the addressee.

Imperatives may thus be thought as having a uniform semantic core of preference, but can be used to convey various speech acts depending on contextual conditions. Chief among those acts is the act of command, which relies on the speaker’s goal to make the addressee bring about action to achieve that goal. In this sense, the command reveals the strongest force of the imperative, since the speaker is personally invested in having their goal realized. The other three types we distinguished are rooted in the perspective of the addressee, hence the speaker is less invested in their realization—they can therefore be seen as weaker from the speaker’s perspective. In other words, we can view the four types in (4) as realizing a two-way distinction based on speaker perspective and strength: speaker-oriented imperatives (command, strong), versus addressee-oriented imperatives (weaker from the speaker’s

perspective). The latter category is the one that involves more variability in illocutionary force, it is therefore not unreasonable to expect more variability in the means of realization.

Non-manual marking of imperative speech acts expresses important pragmatic information that can be used to specify the particular act expressed with an imperative. Although these non-manual markings are not totally conventionalized (cf. Donati et al., 2017), we assume that the different uses of imperatives can be understood by prosodic cues alone. In the three studies presented here, we ask how strongly and how consistently the pragmatic differences, i.e. the illocutionary forces of the imperatives in (4), are encoded in ASL and understood by ASL signers and three other groups of signers and non-signers without exposure to ASL.

With regard to comprehension, we are interested in determining across groups whether the cues for the imperative types show specific groupings—e.g., speaker-(command) vs. addressee-oriented imperatives (permission, advice, explanation), or, “must” type imperatives (command, explanation) vs. “may” type imperatives (permission, advice). With regard to the groups, we entertain two hypotheses, which may seem like they are competing, but we expect the results to support both of them, at least to some extent. The Hypothesis of Universality (Hypothesis A) predicts that the cues marking of pragmatic distinctions in imperatives reflects universal expressive strategies based on facial expressions, such as those described in Ekman and Friesen (1971). If this is the case, the meanings should be accessible to all of the groups in our studies equally, and knowledge of ASL should not provide any advantage. The Hypothesis of Arbitrariness (Hypothesis B) predicts that the cues marking of pragmatic distinctions in imperatives are entirely arbitrary and language-specific, and the meanings should not be accessible to anyone without knowledge of ASL grammar. We expect the results to support both hypotheses to some extent, since we not only expect the facial expressions marking imperatives to be accessible to all groups, but also that their particular grammatical distribution in ASL grammar will offer a significant advantage to ASL signers in distinguishing among imperative types. Study 1 involves the production of five conditions (the four types of imperatives mentioned above and neutral sentences) by a Deaf native ASL signer. We annotate and analyze several different prosodic cues for scope and quantity across the five conditions. In Studies 2-3, we then use these production data as stimuli in a task designed to study the comprehension of the speech acts corresponding to the five conditions. Study 2 examines their comprehension by other Deaf native and early learners of ASL. Study 3 expands the groups of participants performing the comprehension task to include a group of Deaf DGS signers, and two groups of non-signers: a group from the United States and a group from Germany. All three studies were approved and carried out in accordance with the recommendations of the Internal Review Board of the University of Chicago for the ethical treatment of human subjects, and with written informed consent from all subjects (IRB protocol 14-0410). All subjects gave written informed consent in accordance with the Declaration of Helsinki.

²We refer to the semantics/pragmatics interface together because we are concerned with sentence meaning that originates from sentence-internal factors (semantics) or from the surrounding discourse (pragmatics).

STUDY 1

Materials and Methods

Participants

The sentences were produced by one Deaf, third-generation, native ASL signer (male, age 36).

Stimuli

All items consisted of two signs, which were combinations of four verbs consisting of a single path movement—PICK/FIND (these signs are homophonous), TAKE, THROW-AWAY, and KEEP—and four nouns consisting of a 2-movement reduplicated sign—BOOK, HAT, PAPER, and WATCH. Each sentence appeared in five conditions: neutral, command, explanation, permission, and advice (16 sentences \times 5 conditions = 80 items). The number of words and syllables per sentence was therefore uniform across items, as was word order. Verb+Noun_{DO} is the unmarked order for all sentences employed. The neutral clause was extracted from the sentence frame “I LIKE” to ensure a neutral production (i.e., a declarative sentence expressing assertion).

Procedures

Definitions and instructions were given in ASL. The signer was told that the task was about understanding the meaning of ASL sentences. After providing the signer with definitions of the imperative types and examples of contexts in which each of the speech acts would be produced, such as those in (4), he was instructed to construct an imagined context to achieve the targeted imperative type for each item presented. A set of 8–10 practice items were then presented. After the experimenter was satisfied that the signer understood the task and was comfortable with it, the 80 items were each presented in pseudo-random order using a Powerpoint presentation. The signer could control the pace of presentation. The types of imperatives were prompted by a static image of a sign for the five types of sentences—neutral, command, explanation, permission, and advice—followed by static images of the two signs making up the sentence. The signer was allowed to repeat the sentences until he was satisfied with the production for each item. The clips he judged to be the most representative for each item were then clipped and annotated in ELAN (<https://tla.mpi.nl/tools/tla-tools/elan/>). The annotations were completed by a research assistant who is a fluent second language learner of ASL and was trained in annotating non-manual cues. 20% were re-annotated by the first author with reliability of 95% for cues, and 90% for the duration of each sign. After discussion all discrepancies were resolved.

The cues that were analyzed are listed and defined in **Table 1**, and these were annotated for the verb and the noun separately. The manual cues were sign duration and hold duration, and the non-manual cues were head nod, head tilt, mouthing, and eyes wide. In keeping with the distinction we made in the introduction, annotators were sensitive to the possible use of “expressive” and “iconic” non-manuals. We wanted to analyze focus on expressives in this paper, so we constructed stimuli with simple verbs and nouns that were not prone to iconic non-manuals. As predicted, we found no manner, size, or shape non-manuals in the signer’s productions.

The set of cues included in the analysis was arrived by first annotating a much larger set of cues that are associated with intonational phrases and have been observed in the literature (Nespor and Sandler, 1999; Brentari and Crossley, 2002; Pfau and Quer, 2010; Brentari et al., 2011, 2015; Sandler, 2012). In addition to the six cues in **Table 1**, we annotated transition duration between signs, brow raise, brow furrow, body lean, squint, single head nods, smile, and corners of the mouth turned down, but the cues in this last set were used too rarely or showed no relevant pattern, and so were not included in the analysis. We then added cues that we saw in the data that were previously unattested. We added eyes wide to characterize a very open eye position accompanied by a penetrating eye gaze that appeared frequently in these data.

Examples are given in (5) of one sentence across all conditions with its annotated cues; the distribution of cues is presented in the Results section. Sign duration is noted by adjusting the space between the glosses. Since these cues are relative and dynamic, video examples are provided in the Supplementary Materials.

(5) Example ASL sentences (See also Supplementary Materials, Videos 5–9)

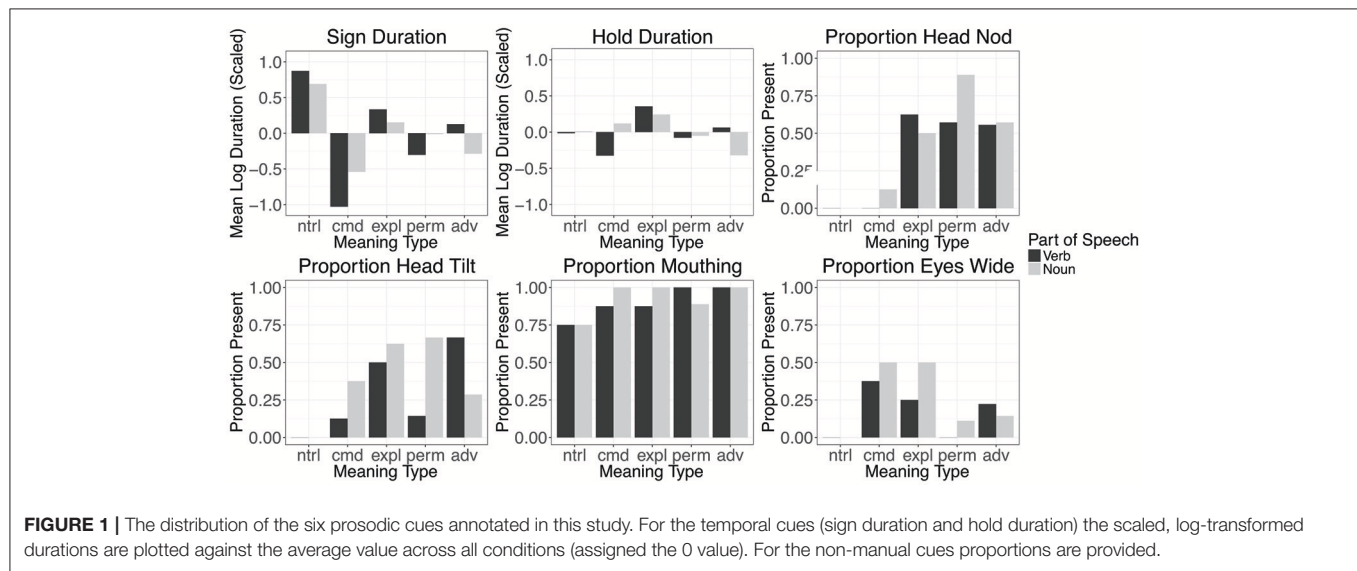
- | | | | |
|-----------------------|------|------|------|
| a. Neutral: | | | |
| mouthing | | | |
| | FIND | | BOOK |
| b. Command: | | | |
| eyes wide | | | |
| mouthing | | | |
| | FIND | BOOK | |
| c. Explanation | | | |
| eyes wide | | | |
| mouthing | | | |
| holds | | | |
| | FIND | | BOOK |
| d. Permission | | | |
| head nod | | | |
| head tilt | | | |
| mouthing | | | |
| | FIND | | BOOK |
| e. Advice | | | |
| head nod | | | |
| head tilt | | | |
| mouthing | | | |
| | FIND | | BOOK |

Results

We analyzed each of the cues with regard to its use on the verb and on the noun in the 80 sentences. The distribution of each cue across conditions is given in **Figure 1**. For the temporal cues (sign duration and hold duration), we applied a log-transformation to the values, and then scaled the log-transformed durations to have a mean of 0 and a standard deviation of 1. We then report the average scaled duration for each meaning type. Thus, values

TABLE 1 | Prosodic cues analyzed in the productions of imperatives.

Property	Definition
1. Sign duration	Length of time from full formation of initial handshape to initial decay of final handshape; this measure includes sign-final holds
2. Hold duration	Periods during which the handshape and location of the sign were static
3. Head nod	Continuous nodding of the head during the production of a sign
4. Head tilt	Tilts of the head backward, forward, or sideways
5. Mouthing	The silent production of some or all of the corresponding English word
6. Eyes wide	Eyes more widely open, accompanied by a penetrating gaze



below 0 indicate shorter durations than the overall average across all conditions, and values above 0 indicate longer durations than the average across all conditions. For the remaining cues, we report the proportion of signs that expressed each cue at any point during the sign.

A summary of our findings is as follows. The average sign duration is longer in neutral sentences than all other conditions, and shorter in commands than in all other sentence types. Hold duration has more modest effects, but sentences of explanation have longer than average holds on both signs, while commands have shorter holds on the verb, and sentences of advice have shorter holds on the nouns. Sentences of explanation, permission, and advice all employ head nod to some degree, and sentences of permission have an increased use of head nods on the noun; in contrast, neutral sentences and commands rarely use this cue. Head tilts also occur with sentences of explanation, permission, and advice more frequently than with neutral sentences or commands; they are more likely to appear on the verb in sentences of advice, on the noun in sentences of permission, and on both in explanations. Mouthings accompany exclusively the verbs and nouns; no other mouthing or mouth gestures occurred. Mouthings occur quite frequently, but are less frequent in neutral sentences. The cue eyes wide appears most frequently in commands, and also appears frequently on the nouns of explanations.

We used a multinomial logistic regression model on the cues to try to predict the condition. We used 4-fold cross-validation

to assess the accuracy of the model. The data are randomly split into 4 segments, and the model is trained on each possible set of three segments and then used to predict the remaining segment. We then compute the accuracy of these predictions. Because the model is predicting across five conditions, a baseline chance performance is 0.20.

Neutral sentences and commands are predicted well above chance, and as presented in **Table 2** they are rarely mistaken for other sentence types. Explanation, permission, and advice sentences are rarely mistaken for other meaning types, but they are frequently mistaken for one other. From among these three types, permission is predicted most accurately and it is mistaken for other meaning types relatively least often, whereas sentences of advice and explanation are often mistaken for one another.

Discussion

From the analyses above we can arrive at several generalizations concerning the distribution of cues. This can be schematized as in (6).

(6) Reliability of prediction of meaning types based on the regression model

Neutral > Command > Permission > Explanation, Advice

Neutral sentences could be identified as distinct from any of the imperative types because they displayed the fewest non-manual

TABLE 2 | Results of the Logistic regression model for Study 1 (Production task).

Stimulus	Neutral	command	Explanation	Permission	Advice
MODEL OUTPUT					
Neutral	0.81	0	0.06	0	0.12
Command	0	0.62	0.19	0.12	0.06
Explanation	0	0.12	0.38	0.25	0.25
Permission	0	0.12	0.25	0.5	0.12
Advice	0	0.06	0.44	0.19	0.31

The numbers in bold indicate accurate responses.

cues, both in type and frequency. Neutral sentences also had longer average sign durations than any of the imperatives. In essence, the lack of non-manual cues and the relatively long durations were rather strong indications that the sentence is a neutral sentence. It has been shown that the presence and the absence of cues may be informative for sentence meaning (Herrmann, 2015).

Of the imperative sentence types, commands could be identified by the non-manual cue of eyes wide, along with shorter sign durations. Commands were also less likely to have head nods, head tilts, and the holds on verbs were shorter than average. Explanations had longer holds on both signs, and sentences of advice had shorter than average holds on the noun, but these effects were relatively modest. The imperative findings in Study 1 suggest clearly a pattern of commands versus the rest, and this divide maps onto the notion of speaker goal vs. addressee goal outlined in the theoretical work on these speech act types.

Addressee goal imperatives such as advice, explanation, and permission do not clearly have unique prosodic patterns. The cues that appear on these sentence types were subtle distinctions of distribution, and sometimes appeared on only one of the two signs—either the verb or the noun. For example, sentences of advice and permission both used head tilts but advice was more likely to have this cue on the verb, and permission more likely to have it on the noun. From among sentences of advice, explanation, and permission, sentences of permission are predicted more reliably than those of advice and explanation. One interpretation of the variability would be that all items without a clear absence or strong presence of prosodic markers are unclassified. Another interpretation is in agreement with the weaker nature of those imperatives, i.e., weaker in the sense that the speaker is less invested in their realization (as noted earlier). Given that the addressee's investment is variable, the observed flexibility is expected.

The regression model provides some predictions about what humans might attend to in evaluating these sentences. With regard to the type of cues, we expected both temporal and non-manual cues to differentiate these meaning types and, indeed, that is what we found. Since our sentences were from one signer, we cannot rule out that other cues may also fill these same roles, or that a pattern of general prominence is also factor, in addition to the specific cues we found here. We now turn to the two studies of comprehension of these cues by a group of signers of ASL (Study 2) and by three other groups (Study 3): DGS signers from Germany, non-signers from the USA, and non-signers from Germany. These studies will help us understand which cues

employed to identify these meaning types are accessible to the different groups.

STUDY 2

Study 1 has informed us about the strength and frequency of a set of six prosodic cues and their patterning in the five target meanings, but they do not address whether the regression model is predictive of the comprehension of these meanings by ASL signers, nor what other cues may influence ASL signer judgments. For example, the degree of tension in the body, and movement acceleration and deceleration, are both noticeable, but we could not reliably annotate these cues from the video stimuli; therefore, the results of this study will help us determine if the cues we annotated are indeed those that signers attend to when identifying these five types of meanings based on prosody.

Materials and Methods

Participants

Thirteen adult, Deaf native or early learners of ASL signers from the United States participated in this study. Eight participants learned ASL from birth, while five participants were early learners who acquired ASL prior to age 5 (eight females; five males). The ages of our participants were as follows: three were 18–25 years, two were 25–35 years, four were 35–45 years, one was 45–55 years, two were 55–65 years, and one was over 65 years.

Stimuli

The stimuli consisted of the 80 sentences analyzed in Study 1. A sample video for each of the meaning types is provided in the Supplemental Materials.

Procedures

Using a web-based interface, participants completed both a multiple choice and a matching task. In this paper, we analyze only the multiple-choice task³. The instructions for the task and definitions of the meaning types and sample contexts were presented in ASL on videotape to ensure consistency across participants, along with some English text to label the conditions and sentence choice options. The participants watched the instructions before proceeding to two practice sentences using commands and neutral sentences. The 80 two-sign sentences were presented in pseudo-random order, and signers were free to return to the definitions and sample contexts as often as needed. The 80 sentences were split evenly between the multiple choice and the matching tasks, with half of the participants completing the multiple choice task first and half of the participants completing the matching task first. For the multiple choice task, each item consisted of a slide containing the video and 5 multiple choice buttons with labels corresponding to the meaning type. Participants were asked to pick the meaning that they thought was being expressed. The items were presented in blocks of 10.

³We thought the matching task might be easier, but there was no statistical difference between types of items, so in Study 3, two of the three groups were given only multiple-choice items. To have comparable data for analysis across all groups we therefore analyzed only the multiple-choice items here.

TABLE 3 | Accuracy and confusion matrices for Study 2 (Comprehension-ASL signer group).

Condition	Neutral	Command	Explanation	Permission	Advice
ASL SIGNERS' RESPONSES					
Neutral	0.93	0.00	0.03	0.02	0.02
Command	0.01	0.82	0.07	0.02	0.09
Explanation	0.06	0.01	0.35	0.24	0.35
Permission	0.02	0.00	0.15	0.50	0.33
Advice	0.05	0.04	0.26	0.32	0.34

The numbers in bold indicate accurate responses.

Results

The accuracy and confusion matrices are provided in **Table 3**. The results are strikingly similar to the predictions of the regression model of Study 1, with comparable confusions among the same imperative types. ASL signers were even better at identifying neutral sentences and commands than the regression model predicted in Study 1 (0.93 and 0.82). Explanation, permission, and advice are about as accurate as would be predicted by the model and confusable in the same ways. Among the group that includes explanation, permission, and advice, sentences of permission are slightly easier to predict (0.50).

Discussion

Because the results from the Deaf native ASL signers are so similar to the regression model results, we can be reasonably certain that we have annotated most of the relevant cues that ASL signers employ to make their judgments of these five meanings. We can also say with some certainty that ASL signers are—as expected—able to identify speech acts on the basis of temporal and non-manual prosodic cues.

STUDY 3

We now turn to the question of the accessibility of the ASL prosodic cues used for imperatives by three additional groups using the same comprehension task as was used in Study 2. The groups are: signers of DGS, non-signers from the United States, and non-signers from Germany. The DGS signers' results will inform us about how accessible the meanings of the prosodic cue patterns are to people without exposure to ASL, but with knowledge of a sign language and who are accustomed to attending to the hands and face for prosodic cues. The results of the American, hearing non-signers inform us about accessibility that might be due to shared gestural competence based on residing in the same country. The results of the German, hearing non-signers will inform us about broader accessibility of these patterns of prosodic cues, at least extending to communities whose origin is Western Europe.

Questions of accessibility also indirectly address the issue of how these cues come to be conventionalized, especially because the imperatives utilize facial expressions of emotions and mental states.

Materials and Methods

Participants

The participants in this study consisted of three groups. Group 1 consisted of fifteen adult, Deaf native or early learners of DGS from the Göttingen area who had no knowledge of ASL. Ten participants learned DGS from birth, and five participants were early learners who acquired DGS prior to age 7 (9 female; 6 males). The ages of our participants were as follows: four were 18–25 years, three were 25–35 years, one was 35–45 years, four were 45–55 years, two were 55–65 years, and one was over 65 years.

Group 2 consisted of 17 hearing American non-signers (recruited through Amazon Mechanical Turk) who had no knowledge of any sign language (7 females; 10 males). The ages of our participants were as follows: three were 18–25 years, nine were 25–35 years, and five were 35–45 years.

Group 3 consisted of 15 German non-signers who had no knowledge of any sign language (5 females; 10 males). The ages of our participants were as follows: four were 18–25 years, and 11 were 25–35 years.

Stimuli

The stimuli consisted of the 80 sentences analyzed in Study 1. Like the American signers, the American non-signers completed both a matching and a multiple-choice task. The German signers and non-signers performed the multiple-choice task for all 80 sentences. A sample video for representative sentences for each of the meaning types is provided in (5) and in the Supplementary Materials.

Procedures

Instructions, definitions and contexts for the American non-signers were translated from ASL into English and presented as English text. The instructions, definitions and contexts for the German non-signers were translated from English into German and presented in German text. The instructions, definitions and contexts for the DGS group were translated from German into DGS, videotaped and presented in DGS with some German text to label the buttons, etc., parallel to the ASL instructions of Study 2. The other procedures for Study 3 were the same as for Study 2.

Results

The accuracy and confusion matrices for all three groups are provided in **Table 4**. There are two main results. First, the ASL signers from Study 2 as a group performed better than the other three groups, and second, the three non-ASL groups performed similarly to both the predictions of the regression model of Study 1, and to the ASL signer comprehension results in Study 2. Like the ASL signers, these three groups were better at identifying neutral sentences and commands, and they had less accuracy and more confusion in identifying sentences of explanation, permission, and advice.

In order to confirm these impressions, we also used a logistic regression model that predicts whether participants gave the correct response on each item. We included meaning type and participant group as predictors, as well as the interaction between these terms (in case some group performs significantly better or

TABLE 4 | Accuracy and confusion matrices for the DGS, American non-signer, and German non-signer groups.

Stimulus	Neutral	Command	Explanation	Permission	Advice
DGS SIGNERS' RESPONSES					
Neutral	0.78	0.05	0.06	0.06	0.04
Command	0.02	0.75	0.11	0.07	0.05
Explanation	0.08	0.03	0.32	0.2	0.37
Permission	0.05	0.06	0.19	0.37	0.33
Advice	0.05	0.12	0.22	0.29	0.32
AMERICAN NON-SIGNERS' RESPONSES					
Neutral	0.78	0.02	0.13	0.03	0.04
Command	0.01	0.76	0.16	0.01	0.05
Explanation	0.06	0.11	0.29	0.24	0.30
Permission	0.04	0.05	0.26	0.44	0.21
Advice	0.05	0.13	0.29	0.26	0.26
GERMAN NON-SIGNERS' RESPONSES					
Neutral	0.83	0.02	0.08	0.04	0.03
Command	0.03	0.67	0.16	0.04	0.11
Explanation	0.08	0.06	0.41	0.18	0.27
Permission	0.05	0.06	0.26	0.35	0.29
Advice	0.06	0.08	0.33	0.27	0.26

The numbers in bold indicate accurate responses.

TABLE 5 | Results of the Logistic regression model for Study 3 (Comprehension task-all groups).

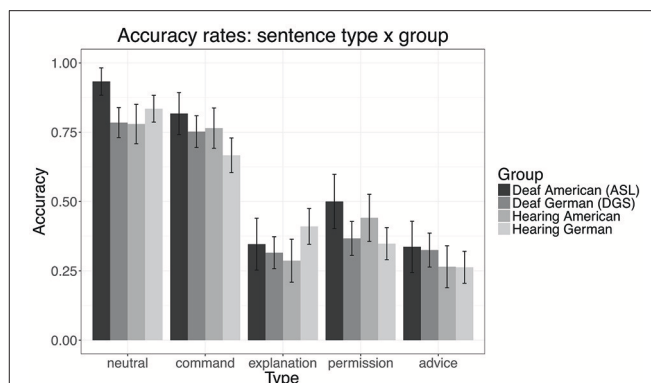
	Estimate	Std. Error	t-value	Pr(> t)
(Intercept)	0.9	0.025	36.06	<2e-16***
Dgs	−0.08	0.024	3.29	0.00099***
German	−0.11	0.025	4.47	8.0e-06***
American	−0.08	0.026	3.02	0.0025**
Command	−0.09	0.024	3.59	0.00034***
Explanation	−0.52	0.027	−19.45	<2e-16***
Permission	−0.43	0.024	−18.24	<2e-16***
Advice	−0.53	0.023	−22.06	<2e-16***
(Interaction) German × explanation	0.15	0.04	3.66	0.00026***

** means ≤ 0.01 ; *** means ≤ 0.001 .

worse on a particular meaning type). We used stepwise regression with the Bayesian information criterion (BIC) to select relevant predictors, since there are a large number of interactions for the size of our data set.

The results of the logistic regression model after variable selection are presented in **Table 5**. Note that the ASL-signing participants and neutral sentences are the baseline encoded in the intercept for the model. Positive coefficients mean better performance than the baseline, whereas negative coefficients mean worse performance than the baseline.

The negative coefficients for all three groups show that they achieve lower accuracy than the ASL-signing participants. There is no statistically discernible difference between the coefficients for each group. Additionally, the only interaction that was selected was the interaction of hearing German group

**FIGURE 2 |** Accuracy rate reported for meaning type by group, with standard error bars.

and explanations, with hearing German participants identifying explanations significantly better than the other groups. Aside from this one difference, this shows that the three groups have a pattern of performance that is largely the same in terms of overall accuracy, as well as their accuracy with regard to the meaning types: neutral sentences were most accurately identified, then commands, and then the other three meaning types, with permission the best identified of these three types for DGS signers and American non-signers, and advice the worst identified type for German non-signers.

All of the sentence types and all of the groups were selected as significant predictors. Commands have a small but significant negative coefficient, showing that participants are slightly less accurate at identifying this type than neutral sentences. Explanation, permission, and advice all have much greater negative coefficients. There is no statistically discernible difference between the coefficients for explanation and advice, suggesting comparable performance on these sentence types. However, the permission coefficient is slightly smaller, suggesting better accuracy for this meaning type.

These patterns can also be seen in the graph below in **Figure 2**, which shows the accuracy rates by sentence type and group. The interval bars represent 95% confidence intervals for the proportion of correct responses.

DISCUSSION: STUDY 3 AND GENERAL DISCUSSION

Non-manual and temporal prosodic cues of a sign language can be used to distinguish certain speech acts, but the patterns for others are highly confusable. Our results across groups show that commands are distinct among imperative types in that they are the most easily identified type. This result is in agreement with our earlier establishing of command as the strongest (thus most proto-typical) speech act, because the speaker is highly invested in the action she wants the addressee to do. We predicted in our earlier discussion, and indeed found here, a pattern distinguishing two classes of imperatives—commands versus the rest—mapping onto speaker goal vs. addressee goal imperatives

(Condoravdi and Lauer, 2012). The results of our study are thus in line with recent analyses of the semantics and pragmatics of imperatives, and the core distinction we drew in section Imperatives.

Addressee goal imperatives such as advice, explanation, and permission were found to not have unique prosodic patterns. The cues that appear on these sentence types were subtle distinctions of distribution, and sometimes appeared on only one of the two signs—either the verb or the noun. For example, advice and permission both used head tilts but sentences of advice were more likely to have this cue on the verb, and sentences of permission were more likely to have it on the noun. The variability is in agreement with the weaker nature of those imperatives, i.e. weaker in the sense that the speaker is less invested in their realization.

Imperative meanings are accessible to people without exposure to ASL; however, as the cues and their distribution have been further conventionalized in ASL, the ASL signers perform better overall. Hence, the results of Studies 2 and 3 provide evidence that certain non-manual and temporal prosodic cues are integrated in the grammatical system of ASL as speech act indicating devices (for the grammaticalization of gestures, see van Loon et al., 2014).

Turning to our two hypotheses from the beginning of the paper we can conclude that both the Hypothesis of Universality and the Hypothesis of Arbitrariness are to some extent supported. The Hypothesis of Universality predicted that non-manual marking of pragmatic distinctions in imperatives reflects universal strategies for expressing mental states. We found that despite the fact that all groups found this task difficult, all were able to perform the task at above chance levels. The Hypothesis of Arbitrariness predicts that non-manual marking of pragmatic distinctions, such as different uses of imperatives, is entirely arbitrary and language-specific, and the meanings should not be accessible to anyone without knowledge of ASL. And, indeed, despite the fact that the content of the prosodic cues was accessible to non-ASL-signers, the additional knowledge of the patterns of conventionalization gave ASL signs a boost in performance that was significant.

Let us first address the similarities in performance across groups. There are at least two possible reasons why neutral sentence and commands are identified most easily, and sentences of explanation, permission, and advice are highly confusable. The first is the system of cue marking. Neutral sentences have the fewest cues and are, as expected, unmarked. Commands are accompanied by the highest number of cues, so their patterns are more structurally distinct from among these five meaning types. Sentences of explanation, permission, and advice have a more complex system of marking, and the differences among the cue patterns for these meanings are more subtle and less consistent; specifically, the same cues are used across all of them to some degree, appear on fewer of the sentences overall, and the differences among these sentences are rather small.

A second possible reason for the similarity in performance across groups is that the content of the prosodic cues is familiar to all groups, at least to some extent. Imperatives are associated with speech acts, which are associated with specific emotional

and mental state facial expressions that accompany them in canonical contexts. Across groups cues involving mental states might include non-manual cues, such as a stern expression for commands, differing degrees of an inviting expression for advice and permission, as well as specific timing cues, such as a slower articulation for explanation. Moreover, commands demand something of the interlocutor and have more negative valence, while the other three imperative types are offering something to the interlocutor and have more positive valence. These results are in line with the assumptions that properties of the context are relevant to specify the speech act performed with an imperative. The interaction of an underspecified imperative sentence mood with specific pragmatic conditions yields the speech act expressed in a specific contextual setting. In this context, the non-manual and temporal prosodic cues seem to function as speech act indicating devices.

Ongoing pilot data by our team involves two follow-up studies using a set of English sentences that parallels the ASL sentences (Brentari et al., 2017) and suggests that some of the same facial expressions are used in English co-speech gesture and in ASL. In preliminary analyses we found that head tilts were used in English sentences of permission and advice, similar to their use in ASL. Some of the temporal cues had parallel realizations as well; for example, from among the imperatives, explanations tended to be longer than any other imperative type in English, and in ASL as we have seen here, perhaps because of their pedagogical nature.

Kuhn and Chemla (2017) and Domaneschi et al. (2017) provide further evidence that hearing non-signers use facial expressions to indicate various speech acts. Kuhn and Chemla (2017) presented non-signers with four emblematic gestures used in American culture combined with facial expressions indicating four conditions. Expression of assertion, wh-question, yes/no question, and command were combined with thumbs up (“good”), thumb pointing (“him”), wrist tap (“time”), and finger rub (“money”) gestures. For example, for the “money” theme, the possible sentences expressing the speech acts were: *It’s expensive.* (assertion), *Pay up!* (command), *How much is it?* (wh-question), *Do you need money?* (yes/no question). Non-signers were able to match the condition with the facial expression at above chance levels. Likewise, Domaneschi et al. (2017) show that Italian speakers associate certain facial expressions with interrogative and directive speech acts. In particular, action units 1 and 4 indicate questions and action units 4 and 5 commands. As opposed to questions and commands, assertions are not marked by facial expressions. Hence, these studies provide evidence that paralinguistic facial expressions may contribute to the understanding of speech acts in spoken languages.

We now turn to possible explanations for the difference between comprehension accuracy in ASL signers vs. the other three groups. Despite the fact that the DGS signers and non-signers can perform this task at above chance levels, the ASL signers were significantly better. Given the result that ASL signers are significantly more accurate on the comprehension task than the other three groups, the ASL signers are more sensitive to the combinatorial properties of these prosodic cues and to their temporal distribution than the other groups. This emphasizes

that the grammar of a language concerns the distribution of forms as much as the content. Dachkovsky et al. (2013) have discussed a number of language-particular differences in the non-manual grammatical markers of ISL and ASL in ways that are relevant here. They outline the very subtle reasons for why ISL and ASL might demonstrate differences in the distribution of language-specific cues, both for phonetic and semantic reasons. For example, the two sign languages produced squints differently phonetically—ISL signers tighten their lower eyelids to produce a narrowed eye aperture, while ASL signers raise the cheeks to accomplish the same result. Semantically, the given-new distinction in the two languages both use squints, but with different frequency, based on how salient or accessible the information is to the interlocutors (Ariel, 2001). ASL signers use squint to mark given information only when that information is very low in “givenness” (low accessibility), while ISL uses it at both low and mid degrees of accessibility.

We acknowledge that our studies have a few weaknesses. One is that there was only one ASL signer for Study 1 and the production results and subsequent items in Studies 2 and 3 were based on his productions. It would be helpful to see whether the results of Studies 2 and 3 are due to the idiolectal cues of one signer or generalizable across signers. Another is that we did not offer alternatives to participants other than neutral and the four imperative types; we might have included a yes/no-question choice, for example. A third weakness is that, even though the 16 sentences appeared in all 5 meaning types and the lexical signs were not sufficient to arrive at the meaning type, the ASL signers knew the signs and they might have been processing the sentences somewhat differently than the other three groups. A follow-up study could rectify all three of these weaknesses by having more signers and additional groups engaged with different tasks—a yes/no, a multiple choice task, and perhaps even a matching task, and instead of ASL signs, also use nonce signs. This work is just a first step along this path.

CONCLUSIONS

The studies presented in this paper have focused on imperative speech acts that were expressed via prosody alone. These prosodic cues signaling emotions and mental states are only partially grammaticalized. Their content is accessible to non-signers to a large extent, while further conventionalization of these cues via their distribution give ASL signers a positive advantage in identifying the imperative speech acts that we investigated. We would argue that using a consistent distribution in alignment, form, and function is an important step in creating a grammatical form. The content of the form may be accessible to non-signers,

but as they become conventionalized, signers become more sensitive to them in a particular systematic distribution.

AUTHOR CONTRIBUTIONS

DB designed the project, initiated the studies, and oversaw all aspects of data collection, analysis, and interpretation. She also wrote the first draft of the manuscript and consolidated the co-authors' contributions into the final manuscript. JF helped design the study, particularly the online task, did all of the statistical modeling, and assisted in the preparation of the manuscript. EV executed translation of tasks and instructions into German, supervised their translation into DGS, collected the data from the German groups, and assisted in the preparation of the manuscript. AH and MS utilized their physical facilities to carry out data collection in Göttingen, Germany. AG, AH, and MS provided their respective expertise in semantics, non-manuals, and sign language grammars, and assisted in the preparation of the manuscript.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation (NSF) Graduate Research Fellowship under Grant DGE1144082 and NSF research grant BCS1227908. This research was also supported by the Sign Language Linguistics Laboratory, the Center for Gesture Sign and Language, and the Neubauer Collegium for Culture and (University of Chicago), as well as the Sign Language Lab (University of Göttingen) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 254142454/GRK 2070.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00770/full#supplementary-material>

Video 1 | ASL assertion.

Video 2 | ASL negation.

Video 3 | ASL coordinate structure.

Video 4 | ASL conditional structure.

Video 5 | ASL neutral.

Video 6 | ASL command.

Video 7 | ASL explanation.

Video 8 | ASL permission.

Video 9 | ASL advice.

REFERENCES

- Ariel, M. (2001). “Accessibility theory: an overview,” in *Text Representation: Linguistic and Psycholinguistic Aspects*, eds T. Sanders, J. Schilperoord, and W. Spooren (Amsterdam; Philadelphia, PA: Benjamins), 29–87.
- Baker-Shenk, C. (1983). *A Microanalysis of the Nonmanual Components of Questions in American Sign Language*. Dissertation, University of California, Berkeley, CA.
- Benitez-Quiroz, F., Wilbur, R. B., and Martínez, A. (2016). The not face: a grammaticalization of facial expressions of emotion. *Cognition* 150, 77–84. doi: 10.1016/j.cognition.2016.02.004
- Bolinger, D. L. (1983). Intonation and Gesture. *Am. Speech* 58, 156–174. doi: 10.2307/455326
- Borràs-Comes, J., Kaland, C., Prieto, P., and, M., Swerts, M. (2014). Audiovisual correlates of Interrogativity: a comparative analysis of Catalan and Dutch. *J. Nonverbal. Behav.* 38, 53–66. doi: 10.1007/s10919-013-0162-0

- Borràs-Comes, J., and Prieto, P. (2011). Seeing tunes. The role of visual gestures in tune interpretation. *J. Lab. Phonol.* 2, 355–380. doi: 10.1515/labphon.2011.013
- Brentari, D. (1998). *A Prosodic Model of Sign Language Phonology*. Cambridge: MIT.
- Brentari, D., and Crossley, L. (2002). Prosody on the hands and face: evidence from American Sign Language. *Sign Lang. Ling.* 5, 105–130. doi: 10.1075/sll.5.2.03bre
- Brentari, D., Falk, J., Lee, J., Chatzikonstantinou, A., and Giannakidou, A. (2017). *The Role of Prosody in the Production and Perception of Imperatives: Data From American Sign Language & English*. Paris: LINGUAE, Institut Jean Nicod.
- Brentari, D., Falk, J., and Wolford, G. (2015). The acquisition of American Sign Language Prosody. *Language* 91, 144–168. doi: 10.1353/lan.2015.0042
- Brentari, D., González, C., Seidl, A., and Wilbur, R. B. (2011). Sensitivity to visual prosodic cues in signers and nonsigners. *Lang. Speech* 54, 49–72. doi: 10.1177/0023830910388011
- Campbell, R. (1997). “Making faces: coextant domains for language and visual cognition,” in *Relations of Language and Thought: The View from Sign Language and Deaf Children*, eds M. Marschark, P. Siple, D. Lillo-Martin, R. Campbell, and V. S. Everhart (New York, NY/Oxford: Oxford University Press), 147–152.
- Campbell, R., Woll, B., Benson, P. J., and Wallace, S. B. (1999). Categorical perception of face actions: their role in sign language and in communicative facial displays. *Q. J. Exp. Psychol. Sect. A Hum. Exp. Psychol.* 52, 67–95. doi: 10.1080/713755802
- Coerts, J. (1992). *Non-Manual Grammatical Markers: An Analysis of Interrogatives, Negations, and Topicalizations in Sign Language of the Netherlands*. Dissertation, University of Amsterdam, Amsterdam.
- Condoravdi, C., and Lauer, S. (2012). “Imperatives: meaning and illocutionary force,” in *Empirical Issues in Syntax and Semantics 9: Papers From the Colloque de Syntaxe et Sémantique à Paris*, ed C. Piñón (Paris: Colloque De Syntaxe Et Sémantique À Paris), 37–58.
- Corina, D. P., Bellugi, U., and Reilly, J. (1999). Neuropsychological studies of linguistic and affective facial expressions in deaf signers. *Lang. Speech* 42, 307–331. doi: 10.1177/00238309990420020801
- Corina, D. P., and Spotswood, N. (2012). “Neurolinguistics,” in *Sign Language: An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 739–762.
- Crasborn, O., and van der Kooij, E. (2013). The phonology of focus in sign language of the Netherlands. *J. Linguist.* 49, 515–565. doi: 10.1017/S0022226713000054
- Dachkovsky, S. (2007). *Linguistic vs. Paralinguistic Intonation in Sign Language*. Nijmegen: Presentation at the Max Planck Institute for Psycholinguistics.
- Dachkovsky, S., Healy, C., and Sandler, W. (2013). Visual intonation in two sign languages. *Phonology* 30, 211–252. doi: 10.1017/S0952675713000122
- Dachkovsky, S., and Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Lang. Speech* 52, 287–314. doi: 10.1177/0023830909103175
- de Vos, C., van der Kooij, E., and Crasborn, O. (2009). Mixed signals: combining linguistic and affective functions of eyebrows in questions in sign language of the Netherlands. *Lang. Speech* 52, 315–339. doi: 10.1177/0023830909103177
- Domaneschi, F., Passarelli, M., and Chiorri, C. (2017). Facial expressions and speech acts: experimental evidence on the role of the upper face as an illocutionary force indicating device in language comprehension. *Cogn. Process.* 18, 285–306. doi: 10.1007/s10339-017-0809-6
- Donati, C., Barbera, G., Branchini, C., Cecchetto, C., and Quer, J. (2017). “Searching for imperatives in European Sign Languages,” in *Imperatives and Directive Strategies*, eds D. van Olmean and S. Heinold (Amsterdam: John Benjamins), 111–155.
- Ekman, P., and Friesen, W. (1971). Constants across cultural in face and emotion. *J. Pers. Soc. Psychol.* 17, 124–129. doi: 10.1037/h0030377
- Emmorey, K. (1999). “Do Signers Gesture?” in *Gesture, Speech, and Sign*, eds L. S. Messing and R. Campbell (New York, NY: Oxford University Press), 133–159.
- Emmorey, K., Bellugi, U., Friederici, A., and Horn, P. (1995). Effects of age of acquisition on grammatical sensitivity: evidence from on-line and off-line tasks. *Appl. Psycholinguist.* 16, 1–23. doi: 10.1017/S0142716400006391
- Goldin-Meadow, S., and Brentari, D. (2017). Gesture, sign, and language: the coming of age of sign language and gesture studies. *Behav. Brain Sci.* 40:e46. doi: 10.1017/S0140525X15001247
- Hellbernd, N., and Sammler, D. (2016). Prosody conveys speaker’s intentions: acoustic cues for speech act perception. *J. Mem. Lang.* 88, 70–86. doi: 10.1016/j.jml.2016.01.001
- Herrmann, A. (2015). The marking of information structure in German Sign Language. *Lingua* 165 (Pt B), 277–297. doi: 10.1016/j.lingua.2015.06.001
- Herrmann, A., and Pendzich, N.-K. (2014). “Nonmanual Gestures in Sign Languages,” in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 2, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressemer (Berlin: de Gruyter Mouton), 2147–2160.
- Iatridou, S. (2008). *De Modo Imperatives. Lecture Notes for a Course at Ealing 6*. Paris: Ecole Normale Supérieure.
- Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research. *Psychol. Bull.* 115, 288–299. doi: 10.1037/0033-2909.115.2.288
- Kaufmann, M. (2012). *Interpreting Imperatives*. New York, NY: Springer.
- Keltner, D., Ekman, P., Gonzaga, G. C., and Beer, J. (2003). “Facial expression of emotion,” in *Handbook of Affective Sciences*, eds K. J. Davidson and K. Scherer (Oxford: Oxford University Press), 415–433.
- Kuhn, J., and Chemla, E. (2017). “Facial expressions and speech acts in non-signers,” in *Poster Presentation at the 6th Meeting of Signed and Spoken Language Linguistics* (Paris: SSSL 6).
- Liddell, S. (1980). *American Sign Language Syntax*. The Hague: Mouton de Gruyter.
- Liddell, S. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.
- Malaia, E., and Wilbur, R. B. (2012). Kinematic signatures of telic and atelic events in ASL predicates. *Lang. Speech* 55, 407–421. doi: 10.1177/0023830911422201
- Malaia, E., Wilbur, R. B., and Milković, M. (2013). Kinematic parameters of signed verbs at the morpho-phonology interface. *J. Speech Lang. Hear. Res.* 56, 1677–1688. doi: 10.1044/1092-4388(2013)12-0257
- McCullough, S., and Emmorey, K. (2009). Categorical perception of affective and linguistic facial expressions. *Cognition* 110, 208–221. doi: 10.1016/j.cognition.2008.11.007
- McCullough, S., Emmorey, K., and Sereno, M. (2005). Neural organization for recognition of grammatical and emotional facial expressions in deaf ASL signers and hearing nonsigners. *Cogn. Brain Res.* 22, 193–203. doi: 10.1016/j.cogbrainres.2004.08.012
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press.
- Meir, I., and Sandler, W. (2008). *A Language in Space: The Story of Israeli Sign Language*. New York, NY: Lawrence Erlbaum.
- Morgan, G., and Woll, B. (2002). “The Development of Complex Sentences in British Sign Language,” in *Directions in Sign Language Acquisition*, eds G. Morgan and B. Woll (Amsterdam: John Benjamins), 255–276.
- Nespor, M., and Sandler, W. (1999). Prosodic phonology in Israeli Sign Language. *Lang. Speech* 42, 143–176. doi: 10.1177/00238309990420020201
- Pendzich, N.-K. (2016). *Lexical Nonmanuals in German Sign Language (DGS). An Empirical and Theoretical Investigation*. Dissertation, University of Goettingen, Goettingen.
- Pfau, R., and Quer, J. (2010). “Nonmanuals: their grammatical and prosodic roles,” in *Sign Languages*, ed D. Brentari (Cambridge: Cambridge University Press), 381–402.
- Poizner, H., Klima, E. S., and Bellugi, U. (1987). *What the Hands Reveal About the Brain*. Cambridge, MA: MIT Press.
- Portner, P. (2005). “The semantics of imperatives within a theory of clause types,” in *Proceedings of Semantics and Linguistic Theory 14*, ed R. B. Young (Ithaca, NY: CLC Publications), 235–252.
- Portner, P. (2007). Imperatives and modals. *Natl. Lang. Semant.* 15, 351–383. doi: 10.1007/s11050-007-9022-y
- Quer, J. (2012). “Negation,” in *Sign Language: An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 626–646.
- Reilly, J. S., and Anderson, D. E. (2002). “Faces: the acquisition of non-manual morphology in ASL,” in *Directions in Sign Language Acquisition*, eds G. Morgan and B. Woll (Amsterdam: John Benjamins), 159–181.
- Reilly, J. S., McIntire, M., and Bellugi, U. (1990). The acquisition of conditionals in American Sign Language: grammaticized facial expressions. *Appl. Psycholinguist.* 11, 369–392. doi: 10.1017/S0142716400009632

- Sandler, W. (1999). Prosody in two natural language modalities. *Lang. Speech* 42, 127–142. doi: 10.1177/00238309990420020101
- Sandler, W. (2009). Symbiotic symbolization by hand and mouth in sign language. *Semiotica* 174, 241–275. doi: 10.1515/semi.2009.035
- Sandler, W. (2010). Prosody and syntax in sign languages. *Trans. Philol. Soc.* 108, 298–328. doi: 10.1111/j.1467-968X.2010.01242.x
- Sandler, W. (2012). “Visual Prosody,” in *Sign Language: An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 55–76.
- Sandler, W., and Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.
- Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., and Pfau, R. (2015). Event representations constrain the structure of language: sign language as a window into universally accessible linguistic biases. *Proc. Natl. Acad. Sci. U.S.A.* 112, 5968–5973. doi: 10.1073/pnas.1423080112
- van Loon, E., Pfau, R., and Steinbach, M. (2014). “The grammaticalization of gestures in sign languages,” in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 2. eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and S. Tessendorf (Berlin: de Gruyter Mouton), 2131–2147.
- Vogt-Svendsen, M. (2001). “A comparison of mouth gestures and mouthings in Norwegian Sign Language (NSL),” in *The Hands are the Head of the Mouth: The Mouth as Articulator in Sign Languages*, eds P. Boyes-Braem and R. Sutton-Spence (Hamburg: Signum-Verlag), 9–40.
- von Fintel, K., and Iatridou, S. (2017). “A modest proposal for imperatives,” in *Modality Across Syntactic Categories*, eds A. Arregui, M. Rivero, and A. P. Salanova (Oxford: Oxford University Press), 288–319.
- Wilbur, R. B. (1999). Stress in ASL: empirical evidence and linguistic issues. *Lang. Speech* 42, 229–250. doi: 10.1177/00238309990420020501
- Wilbur, R. B. (2000). “Phonological and prosodic layering of nonmanuals in American Sign Language,” in *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima*, eds K. Emmorey and H. Lane (Mahwah, NJ: Lawrence Erlbaum Associates), 213–244.
- Wilbur, R. B. (2003). “Modality and the Structure of Language,” in *Oxford Handbook of Deaf Studies, Language, and Education*, Vol. 1, eds M. Marschark and P. E. Spencer (Oxford: Oxford University Press), 332–346.
- Wilbur, R. B. (2009). Effects of varying rate of signing on ASL manual signs and nonmanual markers. *Language and Speech* 52, 245–285. doi: 10.1177/0023830909103174
- Wilbur, R. B. (2011). Nonmanuals, semantic operators, domain marking, and the solution to two outstanding puzzles in ASL. *Sign Lang. Ling.* 14, 148–178. doi: 10.1075/sll.14.1.08wil
- Wilbur, R. B. (2018). “Production of signed utterances,” in *Handbook of Psycholinguistics*, eds E. M. Fernández and H. M. I. Cairns (New York, NY: Wiley/Blackwell), 111–135.
- Woll, B. (2001). “The Sign that Dares to Speak its Name: Echo Phonology in British Sign Language (BSL),” in *The Hands are the Head of the Mouth: The Mouth as Articulator in Sign Languages*, eds P. Boyes-Braem and R. Sutton-Spence (Hamburg: Signum), 87–98.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer, DS, and handling Editor declared their shared affiliation.

Copyright © 2018 Brentari, Falk, Giannakidou, Herrmann, Volk and Steinbach. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Using the Hands to Represent Objects in Space: Gesture as a Substrate for Signed Language Acquisition

Vikki Janke¹ and Chloë R. Marshall^{2*}

¹ English Language and Linguistics, University of Kent, Canterbury, United Kingdom, ² Department of Psychology and Human Development, UCL Institute of Education, London, United Kingdom

OPEN ACCESS

Edited by:

Wendy Sandler,
University of Haifa, Israel

Reviewed by:

Diane Brentari,
University of Chicago, United States
Adam Schembri,
University of Birmingham,
United Kingdom

*Correspondence:

Chloë R. Marshall
c.marshall@ioe.ac.uk

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 30 June 2017

Accepted: 02 November 2017

Published: 20 November 2017

Citation:

Janke V and Marshall CR (2017) Using
the Hands to Represent Objects in
Space: Gesture as a Substrate for
Signed Language Acquisition.
Front. Psychol. 8:2007.
doi: 10.3389/fpsyg.2017.02007

An ongoing issue of interest in second language research concerns what transfers from a speaker's first language to their second. For learners of a sign language, gesture is a potential substrate for transfer. Our study provides a novel test of gestural production by eliciting silent gesture from novices in a controlled environment. We focus on spatial relationships, which in sign languages are represented in a very iconic way using the hands, and which one might therefore predict to be easy for adult learners to acquire. However, a previous study by Marshall and Morgan (2015) revealed that this was only partly the case: in a task that required them to express the relative locations of objects, hearing adult learners of British Sign Language (BSL) could represent objects' locations and orientations correctly, but had difficulty selecting the correct handshapes to represent the objects themselves. If hearing adults are indeed drawing upon their gestural resources when learning sign languages, then their difficulties may have stemmed from their having in manual gesture only a limited repertoire of handshapes to draw upon, or, alternatively, from having too broad a repertoire. If the first hypothesis is correct, the challenge for learners is to extend their handshape repertoire, but if the second is correct, the challenge is instead to narrow down to the handshapes appropriate for that particular sign language. 30 sign-naïve hearing adults were tested on Marshall and Morgan's task. All used some handshapes that were different from those used by native BSL signers and learners, and the set of handshapes used by the group as a whole was larger than that employed by native signers and learners. Our findings suggest that a key challenge when learning to express locative relations might be reducing from a very large set of gestural resources, rather than supplementing a restricted one, in order to converge on the conventionalized classifier system that forms part of the grammar of the language being learned.

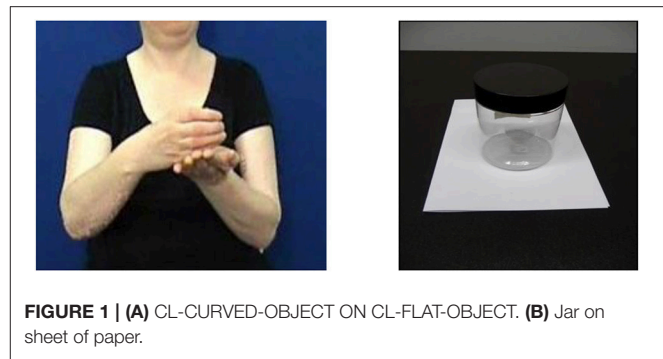
Keywords: gesture, locative expressions, classifier predicates, sign language, sign-naïve adults, adult second language acquisition

INTRODUCTION

This study offers a fine-grained analysis of how adults with no knowledge of sign language (“sign-naïve adults”) begin to use their hands to represent objects in spatial relationships with other objects when required to do so without speech. The relevance of this research potentially extends beyond sign languages to linguistic theory more generally. Any theory of second language acquisition needs to be able to account for data on all languages, including languages in different modalities. An issue of considerable interest in second language acquisition research is what transfers from the speaker’s first language to their second, in other words, identifying specific aspects of cross-linguistic influence (Jarvis and Pavlenko, 2008). Traditionally, one of the reasons that second language learners are thought to differ from native speakers is because their first language “leaks” into the new language. This is evident from foreign accents in pronunciation (Elliott, 2003), from word choice (Carroll, 1992; Janke and Kolokonte, 2015), and from sentence structure (Bardel and Falk, 2007), for example. However, while transfer has been extensively researched in the second language acquisition of spoken languages (e.g., Montrul, 2000; Siegel, 2003; Sharma, 2005; Gabriele, 2009; Gabriel and Kireva, 2014), and even to a certain extent with respect to manual co-speech gesture (Kellerman and Van Hoof, 2003; Gullberg, 2009), it has been largely neglected in studies of sign language acquisition (see Ortega, 2013 for a rare exception).

Having detailed knowledge of what learners start out with in terms of their gestural inventories before they begin to learn a sign language allows us to identify contenders for both negative and positive transfer. It is known, for example, that when asked to reproduce signs that resemble gestures that accompany speech, non-signers bring their gestural knowledge to bear on the task, the result of which can be a less accurately produced sign (see Chen Pichler and Koulidobrova, 2016, for a review of the literature). Conversely, Taub et al. (2008) have identified aspects of some sign novices’ gestures, such as a natural ability to produce handshapes that closely resemble classifiers, which correlate positively with their later ability to engage in third-person discourse in American Sign Language. By providing a detailed picture of gesturers’ manual resources, our study aims to enable connections to be established between what learners produce when acquiring sign and the resources they draw upon.

Our focus is on how objects are represented in space. The visuo-spatial modality of sign languages allows signers to map spatial relationships, such as the relative locations of two or more objects, in a direct and very iconic way using their hands (what Brentari et al., 2012 term “hand-as-object” representations). For example, **Figure 1A** shows a signer of British Sign Language (BSL) expressing the spatial relationship between the two objects in **Figure 1B**. Her handshapes have the meaning of “object from the class of curved entities” (in this particular case, “jar”), and “object from the class of broad and flat entities” (i.e., “sheet of paper”). The orientation of her right hand shows that the jar is upright, rather than on its side or upside down. The location of the curved hand relative to the flat hand shows that the jar is on the paper, and not in any other spatial relationship.



The handshapes that she is using to represent different classes of objects are termed “entity classifiers” (which Zwitserlood, 2012, terms “whole entity classifiers,” and which comprise both what Supalla, 1986, terms “static size and shape specifiers” and “semantic classifiers”; see Schembri, 2003, for a detailed classification of classifiers in sign languages). Importantly, different sign languages do not necessarily choose the same handshapes to represent the same classes of objects; entity classifiers differ cross-linguistically (Frishberg, 1975; Engberg-Pederson, 2010) and the set of handshapes used in classifier constructions is a subset of the handshape inventory for the language as a whole. A speaker would represent relationships such as those in **Figure 1B** very differently, using, depending on their language, lexemes such as prepositions, postpositions, circumpositions, locative case markers, or even positional and posture verbs (see Perniss et al., 2015, for a fuller discussion).

Speakers also make extensive use of their hands during speech (Kendon, 1980; McNeill, 1992), and these gestures complement their verbal communication in interesting ways. Indeed, such co-speech gestures can be similar in form to entity classifier constructions (see Table 1 in Marentette et al., 2016, for a useful summary of the similarities in form between signs and gestures). The frequency with which co-speech gestures occur and the types of gestures that are produced vary according to what a speaker is trying to convey (Kendon, 1980; McNeill, 1992). For example, a study by Lavergne and Kimura (1987) found that conversation involving spatial descriptions elicited double the number of gestures in adults compared to conversation unrelated to spatial descriptions. However, co-speech gestures vary not only in frequency but also in complexity. Representational gestures, for example, contrast with beat gestures (see McNeill, 1992; Kita, 2000; Alibali, 2005), where the former include a heterogeneous set (including handshapes, placement, and movement), which buttress the semantic content of an utterance, and the latter comprise a more basic and limited set of movements, which link to an utterance’s rhythm.

The function of representational gestures means that they might provide the greatest insight into the rich gestural resources that sign-naïve speakers have at their disposal. However, co-speech gestures are rarely used to represent the complete semantic content of the utterances they accompany because this content is already encoded by the spoken words they are associated with. If we are to identify the full extent of the gestural

resources that non-signers can draw upon, we need to provide a context in which the purpose of the handshape they produce is to fully represent a stimulus. In this respect, the term “dedicated gesture,” as introduced by Sandler (2012), is helpful; this describes the gesture recruited for a linguistic purpose, which gradually evolves, reflecting the move from pre-linguistic to linguistic articulation. A first step for researchers, then, toward achieving a closer evaluation of a speaker’s gestural repertoire, is to increase the communicative function of the gesture. This can be achieved by studies of sign-naïve participants in which gestures replace speech altogether, which is the paradigm that we adopt here.

Our examination of how sign-naïve adults gesture visual stimuli that, in signers, elicit classifier constructions is motivated by the gestural properties of these constructions. Although there has long been debate over where classifier constructions are positioned on the gesture-sign continuum (Kendon, 1988; McNeill, 1992; see also chapters in Emmorey, 2003), there is growing recognition that they share many of the properties of gestures. For example, their movement, location, and orientation features are gradient rather than discrete. Furthermore, two-handed classifier constructions are not bound by the linguistic constraints that govern the formation of lexical signs (i.e., the symmetry and dominance conditions identified by Battison, 1978). Indeed, previous studies have shown that hearing adults who are asked to use gestures, but no speech, to describe how objects MOVE in space will produce gestures which have some similarities to sign language classifier constructions (Singleton et al., 1993; Schembri et al., 2005; Brentari et al., 2012). In the current study, we monitor the way in which hearing adults with no knowledge of sign language (“sign-naïve adults”) use their hands to represent STATIC spatial relationships between objects. In particular, we focus on the handshapes that they use to represent the objects that they are locating in space, what Perniss et al. (2015) term “entity representation.” Focusing on static, rather than moving, objects is expected to facilitate greater precision in our comparison of the handshapes of sign-naïve adults and of signers. The depiction of moving objects runs the risk of gesturers choosing to illustrate the path of the movement and not necessarily the object itself (see similar arguments for gestural ambiguity in Alahverdzhieva and Lascarides, 2010). Our focus on static objects avoids this potential confound. It also simplifies the task for participants, who need to concentrate on representing only three parameters, namely handshape, orientation and location, rather than handshape, orientation, and location plus movement.

Our study builds on work by Marshall and Morgan (2015), who investigated how accurately hearing adult LEARNERS of BSL used entity classifier constructions to describe changes in location and/or orientations of objects in pairs of pictures. The learners were all intermediate level students of BSL, who had been learning BSL for between 1 and 3 years. Although classifier constructions have been identified as an area of difficulty for hearing adult learners of sign languages (see Woll, 2012, and references therein), it is not clear which aspects of classifier constructions learners find challenging. Given the transparency of the mapping between the world and the “hand as object” in entity classifier constructions, and their potential gestural origins

(e.g., Okrent, 2002; Liddell, 2003; Schembri et al., 2005), one might predict that they would be acquired easily and therefore produced accurately by this group of learners.

In fact, the learners’ productions did not match those of native signers very well (Marshall and Morgan, 2015). Although they did produce entity classifier constructions on approximately three quarters of the trials, on only one third of trials did they produce a handshape matching that used by native BSL signers. On the remainder of trials they used handshapes which were part of the inventory of BSL handshapes but which were not appropriate for the particular object being represented. Furthermore, learners used handshapes inconsistently, e.g., different handshapes for the same object within a trial. **Figure 2** shows an example of this. The signers are describing a photograph of two people standing next to one another. The native signer, on the left, uses just the index finger to represent “person,” whereas the learner of BSL on the right uses first the index finger and then the flat hand. Marshall and Morgan (2015) also saw learners over-use the flat hand, which replaced other handshapes that native signers used. Thus, there were occasions when learners over-differentiated (i.e., used two or more handshapes to represent the same object) and other occasions when they under-differentiated (i.e., used one handshape to represent two or more classes of object). In contrast to the difficulty with selecting the correct handshape, learners were nearly always accurate at conveying the location and orientation of objects.

This relative difficulty for handshape over location and orientation was only present in Marshall and Morgan’s production task, however. Participants were considerably more accurate in a forced-choice picture-selection task, in which they were presented with trials consisting of four pictures depicting objects in different spatial arrangements (Marshall and Morgan, 2015). Upon viewing the pictures, participants were shown a video-clip of a native signer producing a classifier construction that matched only one of the pictures. Participants succeeded in selecting the matching picture in nearly 90% of trials. Mean error rates for handshape, orientation, and location were all equally low. Importantly, participants did not make more errors for the handshape trials. For example, when presented with a video of a signer signing a classifier construction as in **Figure 1A**, and being shown pictures of different objects on a sheet of paper—jar, apple, coin, and pen—the signers were highly accurate in selecting the

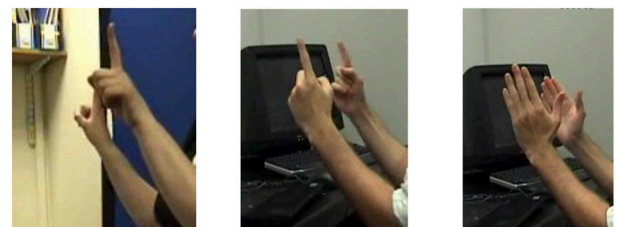


FIGURE 2 | Over-differentiation in a BSL-learner’s representation of two people standing next to one another. The native signer, in the photograph on the left, uses two upright index fingers to represent two people. The learner of sign also uses two index fingers in the photograph in the middle, but then changes to two flat hands in the photograph on the right.

picture of the jar on the paper. In BSL, all those objects would be represented by different classifier handshapes.

Interestingly, when this comprehension task was carried out with a group of hearing adults who had no experience of sign language at all, a similar pattern of success was recorded (Marshall and Morgan, 2015). Although the sign-naïve adults were less accurate overall compared to the learners of sign, they, too, did not make more errors for handshape compared to location and orientation. The fact that they all performed significantly more accurately than chance suggests that it is not particularly challenging for people who have never seen a sign language before to map the shape of a signer's hands onto the correct referent when viewing entity classifier constructions. Note that this is quite unlike spoken second language acquisition, in which one would not expect a person to understand a non-cognate word in a new language on first exposure, thus rendering sign perception unique in this respect.

A question that immediately arises from Marshall and Morgan's (2015) production task is how sign-naïve adults might fare, given that learners found it so much harder to produce classifier constructions than to comprehend them. Because sign-naïve adults will, by definition, bring no sign language experience to the task, their spontaneous creations are likely to build upon their existing gestural abilities (as proposed by, *inter alia*, Taub et al., 2008; Brentari et al., 2012; Ortega, 2013). The learners of sign did have difficulty choosing the correct handshapes to represent the spatial arrangements of objects in the production task. Assuming that gesture is available as a substrate for learning a sign language, there are two alternative possible reasons for this difficulty. (1) Learners might have had few resources from gesture to draw upon, and in particular, a very limited repertoire of handshapes available to them to represent objects. This would imply that they were learning from scratch that the hand can take on different shapes to represent different objects. (2) Another possibility, however, is that learners did in fact have a substantial repertoire of handshapes at their disposal from gesture, and that the difficulties they exhibited in the production task stemmed from their needing to learn to select the appropriate, conventionalized, handshape for each object. On the basis of their participants' patterns, Marshall and Morgan (2015) stated that "gesture provides the substrate or the tools that learners recruit to sign with initially" but also that "this system needs to be reorganized for further development toward the system used by native signers" (p. 78). However, they did not discuss the alternatives (1) and (2), and because they provided no details of which handshapes were produced, it is not possible to tell whether the repertoire of handshapes that the learners were drawing on was smaller or larger than the repertoire of handshapes used by the native signers. This missing piece of the puzzle provides the motivation for the present study. By focusing on the handshapes produced by hearing adults with no experience of sign, we test these two alternatives by asking participants to describe the same pictures as Marshall and Morgan's (2015) participants, using silent gesture. Specifically, we examine how they exploit gesture when attempting to express spatial relationships with their hands. For comparison, we

include a reanalysis of some of the data from learners of sign and native signers in Marshall and Morgan's (2015) study.

If the first alternative is correct—i.e., gesturers have only a limited repertoire of handshapes available to them to represent objects—we predict that participants will use only a limited set of handshapes as they complete the task. Indeed, they might make few attempts to create handshapes to represent objects at all, and might instead rely on pointing and enactments (i.e., positioning their whole body to locate the object) as they attempt to convey locative and distributional information about the objects in the pictures. If the second alternative is nearer the mark—i.e., gesturers have a substantial repertoire of handshapes at their disposal—we should see evidence of creativity with respect to handshapes used to represent objects, which will manifest in participants employing a wide range of handshapes that varies between and within participants. In line with the learners in the previous study, we would also expect to find instances of under- and over-differentiation, as well as some handshapes bearing strong similarities to those used by signers.

METHODS

Participants

Thirty hearing British adults (12 male, 18 female) with a mean age of 32 years (*SD* 14; range 19–62) participated. This was an opportunity sample, drawn from undergraduate and graduate students at the universities where the authors work, and also drawn from the authors' acquaintances. Criteria for inclusion were that participants had never learned a sign language, were native speakers of English and reported no neurocognitive impairments. Confirmation of this information was collected via a brief language-history questionnaire, which was completed at the end of the testing session. Information regarding participants' additional languages was also collected from this questionnaire. 20 out of 30 participants had knowledge of one or more second languages, where knowledge was classified as at least an O-Level or GCSE (or its equivalent) in that language¹. Participants were unaware of the specific research questions and hypotheses of the study—they were merely informed that the researchers were interested in how people use their hands to describe pictures.

In addition to the data from our sign-naïve participants, we reanalyzed for comparison some of the data from the learners of sign and the native signers in Marshall and Morgan's (2015) study. Data from these participants allows us to investigate what the conventionalized classifier system for a sign language (in this case, BSL) looks like, and to determine how close to that system a group of learners has moved. The learners of sign comprised

¹GCSEs (General Certificate of Secondary Education) superseded O-Levels in 1986. The GCSE assessment is taken by students in England and Wales when they are aged 16. It equates roughly with A2 in the Common European Framework of Reference for languages, with students being able to:

- Understand sentences and frequently used expressions related to areas of most immediate relevance.
- Communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters
- Describe in simple terms aspects of their background, immediate environment and matters in areas of immediate need.

12 hearing adults (two male), with a mean age of 28 years (SD 6; range 22–44). They had been learning BSL for between 1 and 3 years. All of them had passed BSL Level 1 (beginner), eight had passed BSL Level 2 (intermediate), and three had begun classes at pre-level 3, in preparation for BSL Level 3 (advanced). Of the four adult native signers (one male), three were deaf and one was hearing.

Procedure

Participants were seated at a laptop in a quiet room and informed that they would see two pictures in succession on the screen. Each picture featured two or more objects, whose location or orientation, or both, changed in the second picture (but the identity of the objects themselves did not change). Objects were chosen to elicit a range of handshapes, and included glasses, pens, books, toothbrushes, and toys such as human figures, airplanes, cars and motorbikes. Picture 1 was presented for 3 s, and then Picture 2 for 3 s, after which participants saw a large question mark on the screen. This was the cue for them to describe the pictures. Specifically, they were asked to explain, using only their hands and no voice, how the two pictures differed from each other, i.e., what had changed. This design had proved very successful at eliciting classifier constructions in Marshall and Morgan's (2015) study—the learners of BSL focused on describing just the relevant aspects of the scene, namely the relative locations and orientations of the objects depicted, rather than describing properties that were irrelevant for our purposes such as the color of the objects and their relative sizes.

The experimenter further told participants that it might help them to imagine that they were explaining the pictures to a profoundly deaf individual (similar to the instructions given by Schembri et al., 2005). Participants were not timed and could control the speed at which they progressed through the task. They were allowed to revisit a trial if they felt unsure about what they had just seen. Their responses were filmed, using a video camera mounted on a tripod, which was situated above and to the left of them if they were right-handed and above and to the right if they were left-handed.

Stimuli

The stimuli were identical to those used in Marshall and Morgan's (2015) study. Two types of construction that elicit

entity classifiers in BSL were included: locatives (i.e., X IS AT Y), and distributive plural forms. The locative construction included three conditions, namely, change of location, change of orientation, and change of both location and orientation (see **Figure 3**). There were 10 trials in each of these conditions. Of the 30 locative trials, three had just one object and six had three objects. The remaining 21 locative trials contained two objects. The distributive plural construction included one condition, namely change of distribution. This condition also contained 10 trials (see **Figure 4**), which resulted in a total of 40 trials for analysis. Two practice items trials were presented immediately after the instructions but not analyzed.

Coding

In order to describe the pictures in the task (see **Figures 3, 4**), native signers divide their description into two parts. They first sign the lexical signs for the objects, and then produce a classifier predicate to give a spatial description of those objects. As reported by Schembri et al. (2005), Brentari et al. (2012), and Brentari et al. (2017), we expected gesturers either to do something similar, i.e., to create gestures to first describe the objects and then to show their relative locations, or alternatively just to describe the locations using gesture (given that their instructions were to “describe what has changed,” and only the location and/or orientation of the objects did change). Like Schembri et al. (2005) and Brentari et al. (2012, 2017), we coded just the spatial description part of the gestural sequence, and within that, only the handshapes that were used for that description, as for the current study we were not interested in the accuracy with which location/orientation were represented.

We anticipated that the set of handshapes produced by sign-naïve gesturers would not map exactly onto the set of conventionalized handshapes of BSL. Our coding system thus needed to capture not only those handshapes that did approximate those made by native signers but also those innovations for which there was no obvious BSL parallel. On this basis, handshapes were coded using the inventory and classification scheme devised for BSL as a whole by Brennan et al. (1984) which identifies five groups of handshapes according to finger joint configuration: fully closed, curved or bent, fingers together, fingers spread, and fingers extended from a closed fist. Every trial was coded by the two authors independently,



FIGURE 3 | Example of a locative trial with a change in both orientation and location.

who discussed any initial disagreements until agreement was reached.

Photographs of all the observed handshapes can be found in the appendix. They fell into four categories:

1. In the BSL inventory, and used by the native signers during the task.
2. In the BSL inventory, but not used by the native signers during the task.
3. Not in the BSL inventory, but variants of handshapes that are.
4. Not in the BSL inventory, and not falling straightforwardly into the five aforementioned handshape classes identified by Brennan et al. (Brennan et al., 1984). We labeled these “miscellaneous.”

We also needed to distinguish between elicited productions that attempted to indicate the location of an object using the hands (analogous to the entity classifiers produced by signers doing this task) and those that indicated location by using the hands or whole body to mimic an action associated with the object or relied on the whole body to represent the object (we coded these as examples of enactment). **Figure 5** illustrates the difference between these two possibilities. In response to the first picture, the participant places her index finger to indicate the position and orientation of a motorbike, which is facing her right. In the second, she relies solely on enactment to depict the change in orientation of the bike, which now faces her left. She uses her hands to represent holding the bike’s handlebars and steering to the left.

Finally, we also coded instances in which a participant pointed to locate an object in space.

Ethical Approval

This study was carried out in accordance with the recommendations of University of Kent’s Research Ethics Advisory Group for Human Participants. All participants gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the University of Kent’s Research Ethics Advisory Group for Human Participants.

RESULTS

In order to investigate the range of handshapes exploited by sign-naïve gesturers in their spatial descriptions of objects, we

carried out three analyses. The first simply calculated how frequently gesturers use their hands to represent the relative location and/or orientation of objects in space. The second focused on the inventory of handshapes that gesturers draw upon in their spatial descriptions, and how that inventory compares to that of native signers and learners of sign. In the third, we investigated whether gesturers consistently produce the same handshape to represent the same object.

The Proportion of Trials for Which Gesturers Used Their Hands to Represent at Least One of the Objects in Their Spatial Descriptions of the Elicitation Stimuli

We first calculated the proportion of trials for which gesturers made at least one attempt to use “hand-as-object.” The total number of trials was 40. The data were negatively skewed, as shown in **Figure 6** below: the group mean was 37.5 (i.e., 93.8%), the median 39 and the mode 40.

Having found that hands were used to represent objects in 93.8% of trials, we then looked to see how participants were responding in the remaining few trials, and found them to be distributed between instances of enactments (2.1%), pointing (3.2%), and trials in which participants failed to attempt to represent the objects at all (0.8%).

Our interpretation of these data is that gesturers readily use their hands to represent objects when describing the relative locations and/or orientations of those objects, as signers do. However, this does not mean that they are doing the same

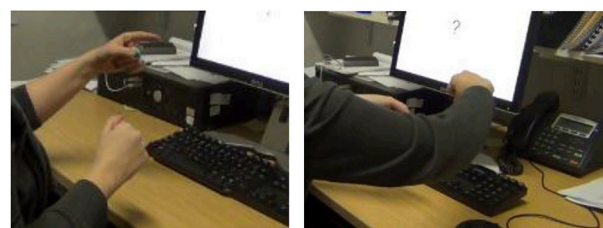


FIGURE 5 | An illustration of the difference between use of an index finger placed horizontally to represent a motor bike (photograph on the left) and enactment of steering a motorbike (photograph on the right).



3s



3s



unlimited time to respond

time →

FIGURE 4 | Example of a distributive trial with a change in distribution.

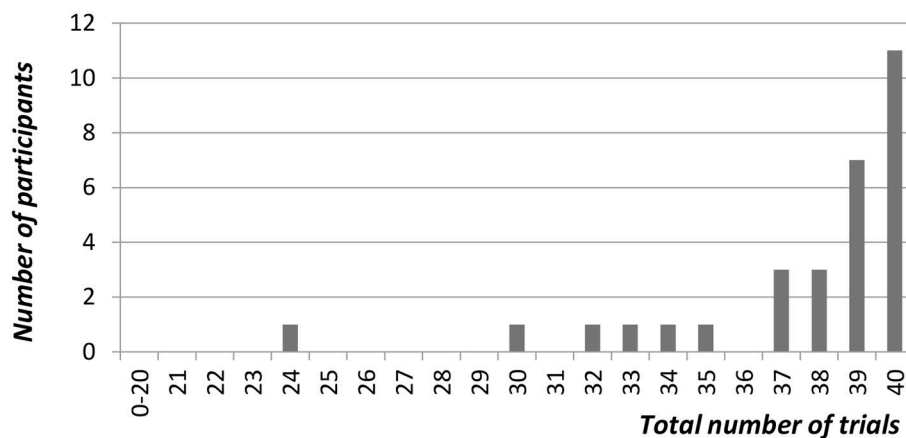


FIGURE 6 | The total number of trials for which participants used their hands to represent an object at least once.

thing with their hands as native signers. Although they use their hands to represent objects, do they create handshapes that are comparable to those of native signers? And how similar are their handshapes to those of hearing adults who are learning sign? We investigate these questions in the next section by comparing the number of handshapes produced by gesturers with those of the native signers and learners of BSL in Marshall and Morgan's (2015) study.

The Inventory of Handshapes That Sign-Naïve Gesturers Drew upon, and How That Inventory Compares to Native Signers/Learners of Sign

We first compared the handshapes produced by gesturers to those produced by the group of native signers. The number of handshapes used by gesturers on the task ranged from 4 to 19 (Mean = 12.47, $SD = 2.99$). From this set, the number of handshapes that overlapped with those used by the group of native signers ranged from 2 to 11 (Mean = 7.10, $SD = 1.99$), while the number of handshapes that were different from those used by native signers also ranged from 2 to 11 (Mean = 5.37, $SD = 2.04$). Therefore, the gesturers each produced handshapes that were the same as those used by signers and likewise each produced handshapes that were different to those used by signers, despite there being a wide range in the number of handshapes that each gesturer used in the task. The only handshape common to all 30 gesturers was the flat handshape (see photograph 16 in the appendix).

Considering the sign-naïve group as a whole, the number of distinct handshapes produced by the sign-naïve gesturers in the task was 53. These handshapes are all shown in the appendix. A reanalysis of Marshall and Morgan's (2015) data revealed that the number of handshapes employed by the native signers and the learners of sign was much smaller, namely 16 for the native signers and 15 for the learners, and those two sets overlapped almost exactly. Gesturers therefore had a very wide selection of handshapes available to them, a near superset of what native

signers and learners of sign used. This situation is illustrated schematically in Figure 7.

From Figure 7, we can see that the sign-naïve gesturers generated all the handshapes that the native signers did. In addition, they created a number of handshapes not found in BSL, or they employed BSL handshapes in different ways to signers. We discuss some relevant observations below, starting with those handshapes not found in BSL.

Firstly, four participants independently converged on a handshape to represent "plane" that does not fit neatly into Brennan et al.'s (1984) classification scheme for BSL, and which to our knowledge does not occur in the inventory of any sign language (although it should be noted that there is no exhaustive list of all the handshapes in existence in all the world's sign languages, so our unawareness of its existence does not imply that it does not exist). This handshape—where the index, middle and ring finger were extended horizontally together, whilst the thumb and little finger were projected out either side—is illustrated in Figure 8, and in photograph 51 in the appendix. Interestingly, Schembri (2001, Table 5.46, p. 229) reported five instances of Australian sign-naïve gesturers using this same handshape when they were describing the motion of an airplane, and noted that none of the Australian and Taiwanese Sign Language users who also took part in the study used it.

One of our participants created an unusual handshape to represent people, which he relied on consistently throughout all trials. His middle, ring and baby fingers were held vertically, pointing down, in what appears to be an attempt to represent a person's legs, whereas his index finger pointed horizontally in the direction in which the person was facing. By exploiting the index finger to encode the direction in which the figure was looking, the participant managed to convey several aspects of his target simultaneously with one handshape. Figure 9 and photographs 48 and 49 in the appendix demonstrate this handshape.

Another interesting observation is that even when gesturers produced a handshape that is part of the BSL handshape inventory, they sometimes used it in different ways to the native signers of BSL and the BSL learners. And yet, what

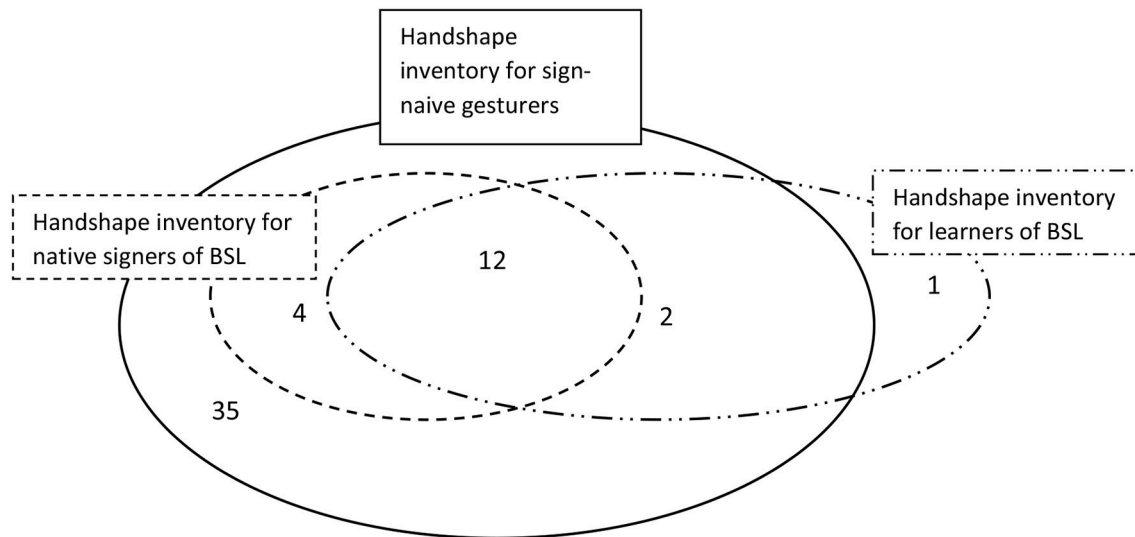


FIGURE 7 | The overlapping entity classifier handshape inventories of sign-naïve gesturers, learners of BSL, and native signers. Numbers represent the number of handshapes at the intersections of the different inventories (e.g., 12 is the number of handshapes that occurs in the inventories of all three participant groups: native signers, learners of BSL and sign-naïve gesturers).

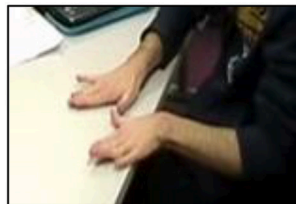


FIGURE 8 | This participant is using a novel handshape to gesture planes lying side by side.

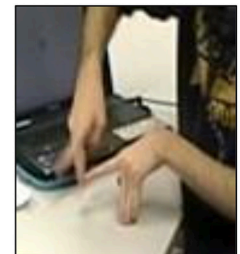


FIGURE 9 | This participant is using a novel handshape to gesture people standing in a row, using the index finger to represent the direction in which the people are facing and the middle, ring and baby fingers the legs.

the gesturers were doing with this handshape has parallels in another sign language. For example, five participants made varied attempts at representing something akin to the “next to” construction found in Turkish Sign Language (but not in BSL). These attempts materialized when participants were faced with photographs in which more than two objects were shown and so had to overcome the problem of having too few hands to depict all the objects simultaneously. Native signers of BSL and learners of BSL overcame this problem by representing objects sequentially rather than trying to represent them all at the same time. Some of our participants, however, constructed a simultaneous representation. One participant, for example, formed a handshape with the index, middle and ring fingers extended (palm down), using the three extended fingers to represent a row consisting of two cars and a person (see Figure 10).

This same participant produced the four handshape when needing to depict four planes, changing the orientation of her hand as shown in Figure 11.

These attempts at expressing the “next to” relation are similar to locative predicates that are licit in Turkish Sign Language

(Özyürek et al., 2010; Perniss et al., 2015). Özyürek et al. (2010), for example, reported an experiment in which six (deaf) Turkish signers were required to describe objects depicted in a photograph in sign to another (deaf) Turkish signer who could not see that photograph. Although the signers used locative predicates far less frequently than classifiers to represent spatial relations, all six of them relied on the locative predicate, “next to” at some point when describing a photograph in which the number of objects was greater than two. The horizontal three handshape, for example, was adopted to depict three plates in a row (see Perniss et al., 2015, Figure 4, p. 621), and the four handshape was produced to illustrate four cups in a row.

In both these examples from Turkish Sign Language, however, the multiple objects that these native signers needed to represent were always of the same type. They were not required to describe the position of several differently shaped objects. For this reason, it is not clear whether the locative predicate would be a licit means of representing different objects (i.e., cars and person,

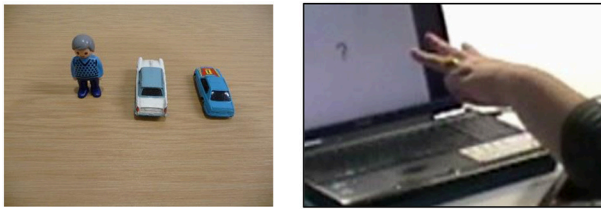


FIGURE 10 | This participant is using the three handshape to represent three objects located next to one another.

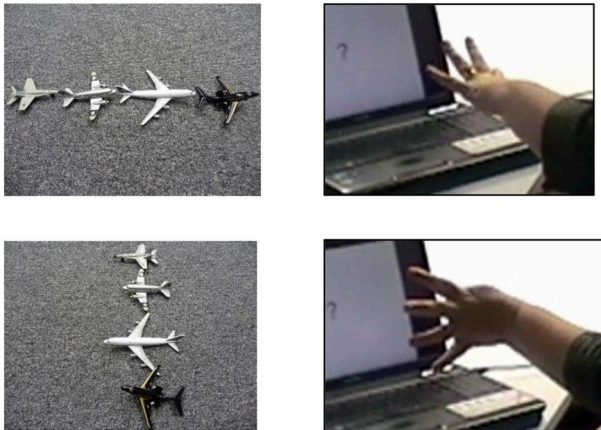


FIGURE 11 | This participant is describing a row of four planes with the four handshape.

Figure 10) next to each other or whether the locative predicate would be restricted to depicting objects of the same type. If the “next to” relation in Turkish Sign Language is restricted in this way, some of our gesturers are showing a more flexible strategy than is permitted in that language. What is interesting for our purposes, however, is that five of our participants came up with this gestural strategy spontaneously when presented with this unforeseen challenge.

Motivated by the same challenge, namely that of depicting more than two objects, some participants converged on other handshapes when attempting to represent these objects simultaneously. Some of the shapes they created can be found in the BSL inventory, albeit with a different function. In one trial, for example, participants were presented with pictures of two cups, one of which held a toothbrush. Three participants employed the handshape illustrated in **Figure 12** below, and in photograph 1 of the appendix, to convey the cup with the toothbrush poking out of it, where their fist depicted the cup and their thumb represented the protruding toothbrush.

In another trial, participants were faced with two pens laying either side of a notepad. One participant made a point shape for one pen but produced a flat hand with her pinkie finger stretched out to the side to capture the second pen (see photograph 53 in the appendix). There were several more examples of these creative efforts to deploy the hands to illustrate three or more



FIGURE 12 | This participant is using two fists to represent the cups, with an extended thumb to represent the toothbrush.

objects simultaneously, seemingly to avoid having to represent them sequentially (such as the handshape in photograph 32 of the appendix, with the index and middle fingers crossed, which some gesturers used to represent two crossed pens so that their other hand was free to represent a sheet of paper).

Despite the large set of handshapes used by the group of gesturers, the majority of them—37 out of 53 handshapes—are licit in the phonological inventory of BSL as a whole (see the appendix), although signers would not use them all in entity classifier constructions. Of the remaining 16 handshapes, seven appeared to be variants of handshapes in the BSL inventory, and four of those occurred only once in our data. The final 9 handshapes do not fit neatly into the classification scheme for BSL handshapes (i.e., the ones labeled “miscellaneous” in the appendix), but only one of those handshapes was used by more than one person (as shown in **Figure 8**). The overall picture with respect to the handshapes used by gesturers can therefore be summarized as follows: (1) At a group level, our sign-naïve gesturers draw on an inventory which is a superset of that used by the native signers in entity classifier constructions, but they rarely produce handshapes that are unlike those found in the entire handshape inventories of BSL and other sign languages; (2) At an individual level, no gesturer draws on exactly the same inventory of handshapes as the native signers of BSL would use in classifier constructions, and they sometimes use handshapes in different ways to the native signers (e.g., by representing two or more objects on the same hand).

We now turn to our third comparison, which considers how consistently participants employed the same handshape to represent a particular object when it occurred several times during the course of the 40-trial experiment.

Handshape Consistency across Trials

The native signers in Marshall and Morgan’s (2015) study proved remarkably consistent in using the same handshape to represent a particular object when that object occurred on multiple occasions. The learners in their study were not so consistent, however, and the sign-naïve gesturers in our current study are even less so. We examined the three groups’ responses for six objects that are represented by different handshapes in BSL, namely car (photo 16 in the Appendix), plane (photo 40), pen (photo 26), book (photo 16), person (photo 26), and glass (photo 7). Each of these objects occurred a minimum of five,

TABLE 1 | Inter-trial consistency of handshapes chosen for six objects by three groups: sign-naïve gesturers, learners of BSL, and native signers^a.

	Sign-naïve gesturers (<i>n</i> = 30)				Learners of BSL (<i>n</i> = 11)				Native signers (<i>n</i> = 4)			
	Number of handshapes				Number of handshapes				Number of handshapes			
	Range	Mean	Median	Mode	Range	Mean	Median	Mode	Range	Mean	Median	Mode
Car ^b	1–9	3.1	3	2	0–3	1.73	2	1	1–2	1.33	1	1
Plane	1–6	2.4	2	2	1–3	1.64	1	1	1–1	1	1	1
Pen	1–6	3.5	3	3	1–3	1.73	2	1	1–1	1	1	1
Book	1–6	2.4	2	2	1–2	1.27	1	1	1–1	1	1	1
Person	1–5	2.9	3	4	0–5	1.18	1	1	1–2	1.33	1	1
Glass	1–6	3.2	3	4	1–3	1.45	1	1	1–2	1.33	1	1

^aData from the latter two groups originate from Marshall and Morgan (2015).

^bAside from car, which occurred six times, there were five occurrences of each object.

and maximum of six, times throughout the task, enabling us to track inter-trial consistency and to compare it across groups. As evident from **Table 1**, which displays the range and central tendency measures of the number of different handshapes used for a particular object, there was most variability in the sign-naïve gesturers, less variability in the learners and least of all in the native-signers.

DISCUSSION

This study investigated the handshapes that hearing adults with no knowledge of sign language create when asked to use just their hands, and no voice, to describe pairs of pictures where the relative location and/or orientation of one or more objects changes. We compared these handshapes with the classifier constructions produced by native signers and by learners of BSL. Hypothesizing that manual gesture is a substrate for sign language learning, we envisaged two potential scenarios: either sign-naïve gesturers would not readily exploit their hands to represent objects when describing their spatial arrangement (and if they did, would produce only a limited set of handshapes relative to signers), or, alternatively, they would employ a much wider set of handshapes than signers. In each instance, there is a gap between the handshapes produced by the silent gesturer (which are not linguistically constrained, e.g., Özçaliskan et al., 2016) and handshapes that a native signer would produce. However, the alternatives diverge in terms of the nature of this gap. Thus, the hypotheses have different implications for the task of the sign language learner.

Three main findings emerged from this study. First, for the vast majority of trials, sign-naïve gesturers used their hands to represent objects and to give spatial descriptions that looked similar in many ways to the entity classifier constructions produced by signers. Second, the group as a whole drew upon an inventory which is a superset of that of those used in the classifier constructions of native signers, and yet they rarely produced handshapes that are unlike those found in BSL as a whole and other sign languages. When looking at individual gesturers, we found that each used some handshapes that were identical to those produced by native signers and some that were not used by native signers, and they sometimes used handshapes

in different ways to the native signers (e.g., by representing two or more objects using a single hand). Third, whereas individual native signers consistently used specific handshapes to represent particular objects, as did learners of sign, the sign-naïve gesturers were much less consistent, with the majority employing a variety of different handshapes to represent the same object across the trials in the task.

We argue that these findings are all consistent with the following interpretation, namely that the challenge for hearing adults when learning to use classifier constructions in a sign language is in learning to select the appropriate, conventionalized, handshapes from a large repertoire of possible handshapes that are available to them by virtue of the large articulatory range of the hands. In other words, the task for learners of sign is not to learn how to represent objects using their hands, but rather to narrow down the set of handshapes that they have potentially available to them to the set of classifier handshapes that is grammatical in the sign language they are learning, and to select from that set accurately and consistently. In the remainder of this section we discuss each of the three findings in turn and motivate our interpretation of the data.

Turning first to the frequency with which gesturers employed their hands to represent objects, we found that the proportion of trials for which gesturers used their hands to represent at least one object on each trial was 94%, which is higher than what Marshall and Morgan (2015) reported for learners of sign (around 75%). We need to be cautious in comparing these figures directly because the task instructions for the two groups were not the same² and so presumably the two sets of participants approached the task differently. Nevertheless, our findings show that learners do not lack the ability to represent objects with their hands, so this cannot be the reason that they find classifier constructions difficult. Gesturers did sometimes draw on other gestural possibilities, such as pointing and enactment (i.e., positioning their whole body to locate the object), but they did so only rarely. Instead, participants appeared to find it quite natural to exploit their hands to represent objects.

²Gesturers were asked to explain, using only their hands and no voice, how the two pictures differed from each other, whereas signers were asked to describe the differences using BSL.

On many trials, gesturers produced handshapes that were not used by the native signers of BSL in Marshall and Morgan's (2015) study, but few handshapes fell outside the repertoire of BSL handshapes, which is consistent with the notion that some handshapes are physiologically harder to produce than others and are hence less likely to occur (Mandel, 1981; Ann, 2006). Interestingly, some gesturers produced handshapes that form part of the inventory of BSL classifier handshapes but they deployed them in ways more akin to other sign languages. For example, some of our gesturers used the handshapes with three and four fingers extended to represent multiple objects lying side by side in a way that is similar to how the "next to" handshape is used in Turkish Sign Language (and possibly in other sign languages too). We also found that objects were not represented consistently, and that speakers often adopted more than one handshape to represent the same object across (and even within) trials. This finding is consistent with the studies of Schembri (2001) and Schembri et al. (2005) for Auslan (Australian Sign Language, which is historically closely related to BSL). Like us, Schembri (2001) and Schembri et al. (2005) investigated sign-naïve participants gesturing silently, although the classifier constructions that they elicited required movement—they were not static like ours. In both studies it was noted that gesturers produced a greater number of handshapes than signers did to represent each category of object. This similarity between Schembri's findings and ours suggests that our task is tapping into resources that are not restricted to the participants who undertook our particular study. Furthermore, assuming that hearing adults draw on the resources available to them in manual gesture when they start learning a sign language, our findings and those of Schembri and his colleagues are consistent with the interpretation that what is challenging for learners of sign is to narrow down the many options provided to them in gesture in order to converge on the narrower conventionalized system of the particular sign language that they are learning.

There are some limitations of our study. The first is an obvious one: sample size. As can be seen in the appendix, only two thirds of handshapes were produced by more than one participant, meaning that the gestural handshape inventory that we compiled would have been different if we had recruited different participants, and likely larger if we had recruited a larger sample. The inventory presented in the appendix is therefore to a certain extent an artifact of sampling, and unlikely to be replicated exactly. Nevertheless, our sample size ($N = 30$) compares well with other similar studies ($N = 22$ in Brentari et al., 2017; $N = 25$ in Schembri et al., 2005), and the study's substantive findings are surely likely to remain if the sample size were bigger.

Secondly, the task was not embedded into a communicative context. A future study might create a paradigm in which the elicited gestures are integrated into a communicative event; such stimuli might give rise to a different pattern of results. A further interesting avenue to explore in subsequent work would be to elicit gestures together with speech for the same items, in order to better understand whether some of the more idiosyncratic handshapes we found are also present in the same

gesturers' co-speech gestural repertoire. We had included in our instructions to participants that "it might help you to imagine that you are explaining the pictures to a profoundly deaf individual," an instruction that has not always been included in previous studies of silent gesture (although it was in the study by Schembri et al., 2005). This instruction might have encouraged participants to create more elaborate handshapes in an attempt to provide greater specificity than would have been the case otherwise. Indeed, there were many examples of participants drawing on iconicity to create handshapes that resembled the form of the objects being depicted much more than the conventionalized BSL handshapes do (recall, for example, **Figure 9**), suggesting that it was important for them to recreate the appearance of objects as accurately as they could. The possibility that our instructions did encourage such elaborate handshapes is not necessarily problematic for our study though—most people who choose to learn a sign language such as BSL do so with the aim of communicating with deaf people, and the learners from Marshall and Morgan's (2015) study (whose data were reanalyzed in the current paper) were presumably approaching this task with communication with deaf people in mind. So although the large number of handshapes elicited in this study may be due to the particular nature of the task itself, and the same participants might produce fewer handshapes in their spontaneous co-speech gesture, we would still argue that we are tapping into speaker's gestural resources. Our particular task has allowed us to uncover just how varied those resources are.

In future work the process of how hearing adults learn a grammatically-constrained classifier system needs to be investigated. So far in our research we have studied gesturers who have never previously been exposed to BSL (current paper) and learners of BSL at a point in time corresponding to between 1 and 3 years of BSL learning (Marshall and Morgan, 2015). We have drawn some inferences about the learning process but, not having studied it directly, we do not know what this process looks like, and, in particular, how the set of classifier handshapes that is grammatical in the language being learned is actually acquired in the early stages of BSL learning. Crucial to such a study would be a close monitoring of the instruction that BSL students receive—in particular, the extent to which classifiers are explicitly taught. One consideration is the relatively low frequency with which classifier constructions occur in spontaneous conversation, as reported in Fenlon et al. (2014), which may have consequences for the time course of linguistic structuring.

Finally, our task could be used as a tool for studying the cultural evolution of sign languages, within the iterated learning paradigm of Kirby and colleagues (Kirby et al., 2014; Motamedi et al., 2017a). In the words of Motamedi et al. (2017b, p. 35), we have investigated in the current study how individuals "improvise solutions to communicative challenges," but we have not yet looked at the next stages in the process of cultural evolution, which are "how groups of individuals create conventions through interaction and how these conventions are transmitted over time through learning." Given the likelihood that sign languages originated as gesture without speech (e.g., Senghas and Coppola, 2001; Sandler et al., 2005), our task is appropriate for determining

a possible gestural inventory available to deaf people when sign languages emerge.

CONCLUDING REMARKS

To understand the nature of second-language learning, it is essential to know what resources learners bring to the task. For adult learners of sign, these resources presumably include manual gesture. Our study aimed to uncover what this manual gesture looks like by eliciting silent gesture with a controlled set of stimuli that in signers elicits entity classifier constructions. We have shown that when sign-naïve adults are required to use silent gesture to describe the locations/orientations of objects, they exhibit a rich repertoire of handshapes. Furthermore, they do not use these handshapes consistently. In contrast, signers have an entity classifier system that is limited to a small set of handshapes which is used in a consistent way. The set of handshapes available to gesturers includes some handshapes that CAN be used as classifiers in the language they are learning and some that CANNOT. Therefore, our findings suggest that the challenge for learners of sign is to narrow down their large repertoire

to the conventionalized system of the particular language they are learning. It remains to be investigated exactly how learners respond to this challenge.

AUTHOR CONTRIBUTIONS

Both authors contributed equally to the work reported in this paper and to the writing of the paper.

ACKNOWLEDGMENTS

We would like to thank all the people who kindly gave up their time to participate in this study, Jessica McMahon for help with the data collection, and Kearsy Cormier for her advice on ways of classifying BSL handshapes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02007/full#supplementary-material>

REFERENCES

- Alahverdzhieva, K., and Lascarides, A. (2010). "Analysing speech and co-speech gesture in constraint-based grammars" in *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar*, ed Stefan Müller (Paris; Stanford, CA: Université Paris Diderot; CSLI Publications), 6–26.
- Alibali, M. (2005). Gesture in spatial cognition: expressing, communicating, and thinking about spatial information. *Spat. Cogn. Comput.* 5, 307–331. doi: 10.1207/s15427633scc0504_2
- Ann, J. (2006). *Frequency of Occurrence and Ease of Articulation of Sign Language Handshapes: The Taiwanese Example*. Washington, DC: Gallaudet University Press.
- Bardel, C., and Falk, Y. (2007). The role of the second language in third language acquisition: the case of Germanic syntax. *Second Lang. Res.* 23, 459–484. doi: 10.1177/0267658307080557
- Battison, R. (1978). *Lexical Borrowing in American Sign Language*. Silver Spring, MD: Linstok Press.
- Brennan, M., Colville, M., and Lawson, L. (1984). *Words in Hand: A Structural Analysis of the Signs of British Sign Language*. Edinburgh: Moray House College of Education.
- Brentari, D., Coppola, M., Cho, P. N., and Sengas, A. (2017). Handshape complexity as a precursor to phonology: variation, emergence, and acquisition. *Lang. Acquis.* 24, 283–306. doi: 10.1080/10489223.2016.1187614
- Brentari, D., Coppola, M., Mazzoni, L., and Goldin-Meadow, S. (2012). When does a system become phonological? Handshape production in gesturers, signers, and homesigners. *Nat. Lang. Ling. Theor.* 30, 1–31. doi: 10.1007/s11049-011-9145-1
- Caroll, S. (1992). On cognates. *Second Lang. Res.* 8, 93–119. doi: 10.1177/026765839200800201
- Chen Pichler, D., and Koulidobrova, H. (2016). "Acquisition of sign language as a second language," in *The Oxford Handbook of Deaf Studies in Language: Research, Policy, and Practice*, eds M. Marschark and P. Spencer (Oxford: Oxford University Press), 218–230.
- Elliott, A. R. (2003). "Phonology: staking out the territory at the turn of the century: integrating phonological theory, research, and the effect of formal instruction on pronunciation in the acquisition of Spanish as a second language," in *Spanish Second Language Acquisition: State of the Science*, eds B. Lafford and R. Salaberry (Washington, DC: Georgetown University Press), 19–45.
- Emmorey, K. (2003). *Perspectives on Classifier Constructions in Signed Languages*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Engberg-Pederson, E. (2010). "Factors that form classifier signs," in *Sign Languages. Cambridge Language Surveys*, ed D. Brentari (Cambridge: Cambridge University Press), 252–283.
- Fenlon, J., Schembri, A., Rentelis, R., Vinson, D., and Cormier, K. (2014). Using conversational data to determine lexical frequency in British Sign Language: the influence of text type. *Lingua* 143, 187–202. doi: 10.1016/j.lingua.2014.02.003
- Frishberg, N. (1975). Arbitrariness and iconicity: historical change in American Sign Language. *Language* 51, 696–719. doi: 10.2307/412894
- Gabriel, C., and Kireva, E. (2014). Prosodic transfer in learner and contact varieties: Speech Rhythm and Intonation of Buenos Aires Spanish and L2 Castilian Spanish Produced by Italian Native Speakers. *Stud. Second Lang. Acquis.* 36, 257–281. doi: 10.1017/S0272263113000740
- Gabriele, A. (2009). Transfer and transition in the SLA of aspect: a Bidirectional Study of Learners of English and Japanese. *Stud. Second Lang. Acquis.* 31, 371–402. doi: 10.1017/S0272263109090342
- Gullberg, M. (2009). Reconstructing verb meaning in a second language. How English speakers of L2 Dutch talk and gesture about placement. *Annu. Rev. Cogn. Linguist.* 7, 222–245. doi: 10.1075/arcl.7.09gul
- Janke, V., and Kolokonte, M. (2015). False cognates: the effect of mismatch in morphological complexity on a backward lexical translation task. *Second Lang. Res.* 31, 137–156. doi: 10.1177/0267658314545836
- Jarvis, S., and Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*. London: Routledge.
- Kellerman, E., and Van Hoof, A. M. (2003). Manual accents. *Int. Rev. Appl. Linguist. Lang. Teach.* 41, 251–226. doi: 10.1515/iral.2003.011
- Kendon, A. (1980). "Gesticulation and speech: two aspects of the process of utterance," in *The Relationship of Verbal and Nonverbal Communication*, ed M. R. Key (Hague: Mouton), 207–227.
- Kendon, A. (1988). "How gestures can become like words," in *Crosscultural Perspectives in Nonverbal Communication*, ed F. Poyatos (Toronto: Hogrefe), 131–141.
- Kirby, S., Griffiths, T., and Smith, K. (2014). Iterated learning and the evolution of language. *Curr. Opin. Neurobiol.* 28C, 108–114. doi: 10.1016/j.comb.2014.07.014
- Kita, S. (2000). "How representational gestures help speaking," in *Language and Gesture*, ed D. McNeill (Cambridge: Cambridge University Press), 162–185.

- Lavergne, J., and Kimura, D. (1987). Hand movement asymmetry during speech: no effect of speaking topic. *Neuropsychologia* 25, 689–693. doi: 10.1016/0028-3932(87)90060-1
- Liddell, S. K. (2003). “Sources of meaning,” in *Perspectives on Classifier Constructions in Sign Languages ASL Classifier Predicates*, ed K. Emmorey (Mahwah, NJ: Lawrence Erlbaum Associates), 199–220.
- Mandel, M. (1981). *Phonotactics and Morphophonology in American Sign Language*. Unpublished doctoral dissertation, University of California.
- Marentette, P., Pettenati, P., Bello, A., and Volterra, V. (2016). Gesture and symbolic representation in Italian and English-speaking Canadian 2-year-olds. *Child Dev.* 87, 944–961. doi: 10.1111/cdev.12523
- Marshall, C. R., and Morgan, G. (2015). From gesture to sign language: conventionalisation of classifier constructions by adult hearing learners of BSL. *Top. Cogn. Sci.* 7, 61–80. doi: 10.1111/tops.12118
- McNeill, D. (1992). *Hand and Mind*. Chicago, IL: University of Chicago Press.
- Montrul, S. (2000). Transitivity alternations in L2 acquisition: toward a modular view of transfer. *Stud. Second Lang. Acquis.* 22, 229–273. doi: 10.1017/S0272263100002047
- Motamedi, Y., Schouwstra, M., Culbertson, J., Smith, K., and Kirby, S. (2017a). “The cultural evolution of complex linguistic constructions in artificial sign languages,” in *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (London: Cognitive Science Society), 2760–2765.
- Motamedi, Y., Schouwstra, M., and Kirby, S. (2017b). An evolutionary approach to sign language emergence: from state to process. [Peer commentary on “Gesture, sign and language: the coming of age of sign language and gesture studies”]. *Behav. Brain Sci.* 40. doi: 10.1017/S0140525X15003003
- Okrent, A. (2002). “A modality-free notion of gesture and how it can help us with the morpheme vs. gesture question in sign language linguistics (or at least give us some criteria to work with,” in *Modality and Structure in Signed and Spoken Languages* eds R. P. Meier, K. Cormier, and D. Quintos-Pozos (Cambridge: Cambridge University Press), 175–198.
- Ortega, G. (2013). *Acquisition of a Signed Phonological System by Hearing Adults: The Role of Sign Structure and Iconicity*, Unpublished doctoral dissertation, University College London.
- Özcaliskan, S., Lucero, C., and Goldin-Meadow, S. (2016). Does language shape silent gesture? *Cognition* 148, 10–18. doi: 10.1016/j.cognition.2015.12.001
- Özyürek, A., Zwitserlood, I., and Perniss, P. (2010). Locative expressions in signed languages: a view from Turkish Sign Language (TİD). *Linguistics* 48, 1111–1145. doi: 10.1515/ling.2010.036
- Perniss, P., Zwitserlood, I., and Özyürek, A. (2015). Does space structure spatial language? A comparison of spatial expression across sign languages. *Language* 91, 611–641. doi: 10.1353/lan.2015.0041
- Sandler, W. (2012). Dedicated gestures and the emergence of sign language. *Gesture* 12, 265–307. doi: 10.1075/gest.12.3.01san
- Sandler, W., Meir, I., Padden, C., and Aronoff, M. (2005). The emergence of grammar: systematic structure in a new language. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2661–2665. doi: 10.1073/pnas.0405448102
- Schembri, A. (2001). *Issues in the Analysis of Polycomponential Verbs in Australian Sign Language (Auslan)*. Unpublished doctoral dissertation, University of Sydney, Australia.
- Schembri, A. (2003). “Rethinking “classifiers,” in *Perspectives on Classifier Constructions in Sign Languages Signed Languages*, ed K. Emmorey (Mahwah, NJ: Lawrence Erlbaum Associates), 3–34.
- Schembri, A., Jones, C., and Burnham, D. (2005). Comparing action gestures and classifier verbs of motion: evidence from Australian Sign Language, Taiwan Sign Language, and non-signers’ gestures without speech. *J. Deaf. Stud. Deaf. Educ.* 10, 272–290. doi: 10.1093/deafened/eni029
- Senghas, A., and Coppola, M. (2001). Children creating language: how Nicaraguan sign language acquired a spatial grammar. *Psychol. Sci.* 12, 323–238. doi: 10.1111/1467-9280.00359
- Sharma, D. (2005). Language transfer and discourse universals in Indian English article use. *Stud. Second Lang. Acquis.* 27, 535–566. doi: 10.1017/S0272263105050242
- Siegel, J. (2003). Substrate influence in creoles and the role of transfer in second language acquisition. *Stud. Second Lang. Acquis.* 25, 185–209. doi: 10.1017/S0272263103000093
- Singleton, J. L., Morford, J. P., and Goldin-Meadow, S. (1993). Once is not enough: standards of well-formedness in manual communication created over three different timespans. *Language* 69, 683–715. doi: 10.2307/416883
- Supalla, T. (1986). “The classifier system in American Sign Language,” in *Noun Classification and Categorization*, ed C. Craig (Philadelphia, PA: John Benjamins), 181–214.
- Taub, S., Galvan, D., Piñnar, P., and Mather, S. (2008). “Gesture and ASL L2 Acquisition,” in *Sign Languages, Spinning and Unravelling the Past, Present and Future. TISLR9, Forty-Five Papers and Posters from the 9th Theoretical Issues in Sign Language Research Conference*, ed R. Müller de Quadros (Petrópolis: Arara Azul), 639–651.
- Woll, B. (2012). *Second Language Acquisition of Sign Language. The Encyclopedia of Applied Linguistics*. Chapple: Blackwell Publishing Ltd.
- Zwitserlood, I. (2012). “Classifiers,” in *Sign Language: An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin: Mouton de Gruyter), 158–186.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Janke and Marshall. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Learning an Embodied Visual Language: Four Imitation Strategies Available to Sign Learners

Aaron Shield^{1*} and Richard P. Meier²

¹ Speech Pathology and Audiology, Miami University, Oxford, OH, United States, ² Linguistics, University of Texas at Austin, Austin, TX, United States

OPEN ACCESS

Edited by:

Marianne Gullberg,
Lund University, Sweden

Reviewed by:

Bencie Woll,
University College London,
United Kingdom
Olga Capirci,
Istituto di Scienze e Tecnologie della
Cognizione (ISTC), Italy

*Correspondence:

Aaron Shield
shielda@miamioh.edu

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 15 December 2017

Accepted: 07 May 2018

Published: 30 May 2018

Citation:

Shield A and Meier RP (2018) Learning
an Embodied Visual Language: Four
Imitation Strategies Available to Sign
Learners. *Front. Psychol.* 9:811.
doi: 10.3389/fpsyg.2018.00811

The parts of the body that are used to produce and perceive signed languages (the hands, face, and visual system) differ from those used to produce and perceive spoken languages (the vocal tract and auditory system). In this paper we address two factors that have important consequences for sign language acquisition. First, there are three types of lexical signs: one-handed, two-handed symmetrical, and two-handed asymmetrical. Natural variation in hand dominance in the population leads to varied input to children learning sign. Children must learn that signs are not specified for the right or left hand but for dominant and non-dominant. Second, we posit that children have at least four imitation strategies available for imitating signs: anatomical (*Activate the same muscles as the sign model*), which could lead learners to inappropriately use their non-dominant hand; mirroring (*Produce a mirror image of the modeled sign*), which could lead learners to produce lateral movement reversal errors or to use the non-dominant hand; visual matching (*Reproduce what you see from your perspective*), which could lead learners to produce inward-outward movement and palm orientation reversals; and reversing (*Reproduce what the sign model would see from his/her perspective*). This last strategy is the only one that always yields correct phonological forms in signed languages. To test our hypotheses, we turn to evidence from typical and atypical hearing and deaf children as well as from typical adults; the data come from studies of both sign acquisition and gesture imitation. Specifically, we posit that all children initially use a visual matching strategy but typical children switch to a mirroring strategy sometime in the second year of life; typical adults tend to use a mirroring strategy in learning signs and imitating gestures. By contrast, children and adults with autism spectrum disorder (ASD) appear to use the visual matching strategy well into childhood or even adulthood. Finally, we present evidence that sign language exposure changes how adults imitate gestures, switching from a mirroring strategy to the correct reversal strategy. These four strategies for imitation do not exist in speech and as such constitute a unique problem for research in language acquisition.

Keywords: sign language, Autism Spectrum Disorders (ASD), imitation, language acquisition, visual perspective-taking, American Sign Language (ASL)

LEARNING AN EMBODIED VISUAL LANGUAGE: FOUR IMITATION STRATEGIES AVAILABLE TO SIGN LEARNERS

Nearly 60 years of research into the signed languages of the Deaf have unequivocally demonstrated that they are fully comparable to spoken languages in a linguistic and biological sense, utilizing similar brain tissue as spoken languages and organized on the phonological, morphological, semantic, syntactic, and discourse levels (e.g., Klima and Bellugi, 1979; Poizner et al., 1990; Emmorey, 2002; Sandler and Lillo-Martin, 2006). They are acquired naturally by children who are exposed to them and achieve language milestones at similar ages as children acquiring spoken languages (Newport and Meier, 1985), and exist as naturally-occurring, autonomous linguistic systems throughout the world wherever a Deaf community is found. Yet signed and spoken languages are not the same. In recent years, many scholars have investigated the role that modality—the channel through which language is expressed and perceived—plays in linguistic structure, highlighting the ways in which signed and spoken languages may differ (Meier et al., 2002). In this paper we focus on the visual-gestural modality of sign in order to identify a crucial difference between the acquisition of sign and speech. We begin with a discussion of the mental representation of lexical signs as a way to frame the unique challenges entailed in sign acquisition.

How do we represent signs? We can represent them as they are typically (but not invariably) viewed by an addressee; that is, from a viewpoint opposite the signer. This is the representation most often seen in videos or in linguistics papers, where photos or line drawings show a frontal view, from waist to head, of a signer. The sign *BLACK*¹ (Figure 1A) viewed from this perspective, moves to the addressee's left (assuming that the signer is right-handed). Yet movement to the addressee's left is not linguistically significant; if the addressee happens to be seated in the passenger seat beside the signing driver of a (left-hand drive) car, the sign *BLACK* moves to the addressee's right. The addressee's usual perspective, opposite the signer, is familiar but is not the basis for a linguistically correct description of the sign.

A better linguistic characterization is that the sign *BLACK* moves to the signer's right, but even this is not quite correct because this description does not capture the way in which left-handed people sign (or even the way in which a right-handed person signs when using the left hand). A still better description is this: in the sign *BLACK* the active hand moves laterally; the direction of movement is away from the signer's midline (and away from the side of the signer that is ipsilateral to the active hand). To take another example, the sign *GIRL* (Figure 1B) makes contact on the signer's cheek, specifically on the cheek that is ipsilateral to the signer's dominant hand. Years of linguistic research have demonstrated that the best description of signs is from the signer's perspective, not the addressee's. Thus, the way in which we generally picture signs does not match the way in which they should be linguistically represented.

¹In this paper we refer to signs from American Sign Language (ASL). However, the hypotheses advanced here apply to other sign languages. As is conventional in the literature, words in SMALL CAPS indicate signs.

Which perspective to take when representing signs has been an issue in attempts to design writing systems for signed languages. In the development of SignWriting, Deaf users instigated a shift from writing signs from the viewer's perspective to the writing of signs from the signer's perspective (Hoffmann-Dilloway, 2017). But, visual representations of signs from the signer's perspective are somewhat unfamiliar; native signers are less accurate and slower to recognize signer-perspective videos of signs than they are to recognize addressee-perspective videos (Emmorey et al., 2009). Maxwell (1980) detected a similar problem in how deaf children decoded drawings of signs in relation to English words (Sign Print). She noted that English print is represented from left to right, but the direction of movement depicted in the drawings of some signs (such as the Signing Exact English plural noun *THING-S*) was from right to left. As a result, a 48-month-old deaf child misinterpreted the signs as occurring in the reverse order (as *S-THING*). She also sometimes turned her body so as to share the same orientation as the figure depicted in the book, evidence of the difficulty posed by the illustrations.

These difficulties in correctly representing signs are not just a problem for linguists seeking to understand grammatical descriptions of a signed language or for people interested in representing signs in written form (or for children attempting to read Signing Exact English). They are a fundamental challenge to children and adults who are acquiring a signed language. We posit that these difficulties present problems for acquisition that are unlike the challenges of acquiring speech. The parts of the body that are used to produce and perceive signed languages (the hands, face, and visual system) obviously differ from those used to produce and perceive spoken languages (the vocal tract and auditory system). In this paper we specifically argue that the asymmetric control of the articulators (the dominant and non-dominant hands) that are used to produce signs, the characteristics of the sign language grammar and lexicon, and the multiple strategies available for the imitation of signs have important consequences for language acquisition and processing. We address each of these issues in turn, marshaling evidence from development, second-language acquisition, and atypical learners to support our observations.

Handedness and the Sign Lexicon

It is perhaps a trivial statement to note that signed languages are produced with the hands, but a few observations about this fact are in order. First, the articulators are paired; under normal circumstances we have two hands. There is no obvious parallel in spoken languages. There are two lips, but they are not involved in the production of every phoneme. Furthermore, the lips are paired vertically rather than horizontally; the same is true for the top and bottom teeth. We would have to imagine a creature with two mouths in a horizontal configuration, each of which could articulate semi-independently from the other, to obtain an adequate analog.

A second observation is that the hands are controlled semi-independently and show different phonological properties. Signs can be one-handed (as in *GIRL*, Figure 1B), two-handed and symmetrical, in which both hands exhibit the same handshape and movement (as in *SCHOOL*, Figure 1B), or two-handed



FIGURE 1 | The ASL signs **BLACK** (A) and the signs **GIRL**, **SCHOOL**, and **TURTLE** (B). All photographs, copyright Aaron Shield and Richard P. Meier, are reproduced here and in **Table 1** with written permission of the model.

and asymmetrical, in which the hands can exhibit different handshapes and in which the non-dominant hand can be static whereas the dominant hand moves (as in **TURTLE**, **Figure 1B**). Battison (1978) discussed the constraints that hold on these three classes of signs in ASL and other signed languages. In the case of one-handed signs, the signer can choose either hand to produce the sign, as signs are not typically specified for left or right². Signers typically employ their dominant hand to produce such signs, though in certain circumstances they may choose to use their non-dominant hand (such as if the dominant hand is holding an object or is otherwise unavailable). The same is true for two-handed asymmetrical signs: the dominant hand acts upon the non-dominant hand, but whether the dominant hand is right or left depends on the individual.

Now let us imagine being a young child who is exposed to a sign language. Handedness is consistently evident by 6 months of age (Butterworth and Hopkins, 1993) or even *in utero* (MacNeilage, 2008), long before children produce their first signs. Let us further imagine that this hypothetical child is an emergent lefty. But all of the adults around him are right-handed, and all he sees is right-dominant signing. This situation must be frequent: it is commonly accepted that about 90% of the general population is right-handed (Corballis, 1980, 1992), and the deaf, signing population shows similar percentages of right-dominance (Conrad, 1979; Bonvillian et al., 1982; Sharma, 2014; Papadatou-Pastou and Sáfár, 2016). How does our imagined child come to understand that he may in fact perform signs with his left hand, when all he sees are examples of right-handed

input? Does his strong motor preference for the left dictate his signing, or does a desire to imitate the exact movements of the adults around him motivate him? We cannot know, of course, what the child is thinking, but we can observe whether the child signs with his right or left hand. In this paper we describe several competing imitation strategies that are available to sign learners and hypothesize that different groups of signers may opt for different strategies due to how they interpret the sign imitation/learning task.

Recent work also suggests that handedness plays a role in sign recognition: Watkins and Thompson (2017) found that left- and right-handed adult signers reacted differently to signs produced by left- and right-handed sign models. Left-handed adult signers responded more quickly on a picture-sign matching task when they viewed two-handed asymmetrical signs produced by a left-handed model than by a right-handed model, suggesting that the articulatory and perceptual complexity of this sign type is more easily recognized when there is congruency between signer and addressee, perhaps because the addressee can more easily recognize the sign through simulation of the sign through their own motor system. However, Watkins and Thompson also found that for all other sign types (i.e., one-handed signs and two-handed symmetrical signs), both left- and right-handed participants identified signs produced by right-handed models more quickly, suggesting a familiarity effect, since both left- and right-handed signers are exposed to more right-handed signing than left-handed signing. Similarly, Sharma (2014) found that left-handed signers made fewer errors than right-handed signers when forced to produce signs with their non-dominant hand, either because they have more practice viewing and processing signs with non-matched handedness or due to weaker handedness than right-handed signers. There is simply no analog to this situation in speech: there is no anatomical component of the vocal tract that varies in such a significant way in a subgroup

²Notable exceptions in many signed languages are signs indicating cardinal directions (EAST is typically produced with a rightward direction and WEST with a leftward direction, in correspondence with the directions on a compass); the same is true for the lexical signs RIGHT and LEFT. We thank one of the two reviewers for this observation.

of the population and that could have such important effects on both production and comprehension of language as does hand preference in sign³.

Perspective-Taking

Like hand preference, the role of visual perspective-taking in sign learning represents a unique challenge for sign learners. Many scholars have noted the role that the three-dimensional signing space plays in sign grammar: the physical space in front of the signer is exploited for pronominal and anaphoric reference, for verb agreement, and for the description of spatial arrays (Bellugi et al., 1990). As far as we know, signed languages universally depict such constructions from the perspective of the signer. Courtin and Melot (1998: 85) point out that such constructions require “a visual perspective change; the addressee has to reorient the linguistic space according to the angle existing between himself and the signer.” Particularly difficult are descriptions of spatial arrays (that is, signed depictions of spatial configurations or movements), as discussed by Emmorey et al. (1998, 2000). Emmorey (2002) provides a schematic for how such constructions are produced and understood. In **Figure 2**, the arrow indicates the direction of movement of a referent, which is represented by the X and is first located in the sign space in front of the signer. The signer wishes to communicate that the referent moved through space, first to the left, then forward, and finally back to the right. The direction of movement is only properly repeated if reversed by the addressee (as in **Figure 2A**), whereas mirroring the movement (as in **Figure 2B**) suggests an incorrect interpretation (right, forward, left). Bellugi et al. (1990: 287) note the challenges that such structures pose to learners: “The young deaf child is faced with the dual task in sign language of spatial perception, memory, and spatial transformations on the one hand, and processing grammatical structure on the other, all in one and the same visual event.” Unsurprisingly, linguistic structures in sign that crucially depend on such mental transformations appear later in development than might otherwise be expected (Lillo-Martin et al., 1985; Newport and Meier, 1985), since “the young deaf child, unlike his or her hearing counterpart, must acquire non-language spatial capacities that serve as prerequisites to the linguistic use of space” (Bellugi et al., 1990: 287).

Shield (2010) proposed that visual perspective-taking and spatial transformations are necessary not only for comprehending complex descriptions of spatial arrays but also for acquiring the phonological form of individual lexical signs. Lexical signs are acquired much earlier in development than are complex spatial descriptions. Unlike the spatial descriptions described by Emmorey (2002), lexical signs are not (typically) specified for right or left, but for the dominant and non-dominant hand. This difference in the use of space has important consequences for the sign-learning child, who must realize that the use of space in lexical signs is fixed and does not

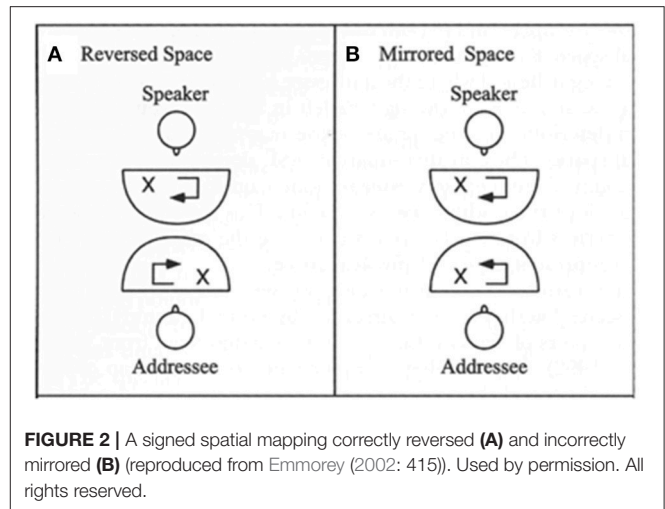


FIGURE 2 | A signed spatial mapping correctly reversed (**A**) and incorrectly mirrored (**B**) (reproduced from Emmorey (2002: 415)). Used by permission. All rights reserved.

make reference to space itself, whereas the descriptions of spatial arrays discussed by Emmorey are linguistic devices for talking about space.

Shield (2010) suggested that some signs engage perspective-taking skills in more challenging ways than others, and that certain types of learners would produce specific error types in the process of learning and reproducing these signs, especially very young typically-developing children as well as individuals with autism spectrum disorder (ASD), who may have difficulties with visual perspective-taking (Hamilton et al., 2009). With regard to the sign types that may be challenging, Shield hypothesized that lexical signs exhibiting *lateral path movements* (from the ipsilateral side of the body to the contralateral side of the body or vice versa) as well as *inward-outward movements* (movements originating at a point distal from the signer's body and moving to a point more proximal to the signer's body or vice versa) require learners to engage perspective-taking skills in order to form correct phonological representations of signs in ways that other types of path movement (such as vertical movements in an upward or downward direction) do not. Likewise, he argued that *inward-outward palm orientations*, such as those found in the ASL signs TUESDAY and BATHROOM (**Figure 3**) could also engage perspective-taking, because these palm orientation values appear differently from the signer's and viewer's perspectives.

Imitation

Like children acquiring speech, children learning sign must imitate the linguistic symbols produced by the language models around them⁴. However, unlike the learning of spoken words, we hypothesize that there are multiple strategies for the imitation of lexical signs, and that not all of these strategies will result in a correctly-formed sign. The psychological literature has distinguished two kinds of imitation strategies: *anatomical* imitation and *mirror* imitation (Koski, 2003; Franz et al., 2007; Press et al., 2009). In anatomical imitation the imitator activates

³An imperfect analogy is perhaps the different vocal tract sizes and resulting fundamental frequencies of male and female speakers (Peterson and Barney, 1952). Six-month-old hearing infants are able to recognize phonemes produced by speakers despite great acoustic variation (*talker normalization*; Kuhl, 1979).

⁴We do not suggest that imitation is the only mechanism through which children acquire linguistic symbols. We focus here on imitation as one essential component of the language-learning process.



FIGURE 3 | The ASL signs TUESDAY (left) and BATHROOM (right). Photographs are reproduced with written permission of the model.

the same muscles as the model being imitated, such that, for example, he raises his right arm to imitate the model's raised right arm, or his left arm to imitate the model's raised left arm. In mirror (or *specular*) imitation, the imitator performs the action as if looking in a mirror (e.g., raising his left arm to mimic the model's lifted right arm). Pierpaoli et al. (2014) found that adults tend to spontaneously engage a mirror imitation strategy more often than an anatomical strategy unless given specific instructions about which limb to use, suggesting that the mirror strategy is a default imitation strategy for typical adults.

For sign-learning children, however, neither the anatomical strategy nor the mirror strategy is correct, for two reasons. First, the anatomical strategy is inappropriate for learners imitating a model with different hand dominance: signs are not specified for the right or left hand but for dominant and non-dominant. Thus the anatomical strategy will fail when the signer and the learner are discordant in hand dominance.

Second, though some signs can be mirrored without error (e.g., signs exhibiting inward–outward movements and palm orientations as in Figure 3 above), the mirroring of signs containing lateral movements (as in the sign BLACK, Figure 1A) will lead to movement reversal errors if signer and learner are both right-handed or both left-handed. In this case, only a *reversing* strategy will result in the production of the correct form. We follow Emmorey (2002) in using the term “reversing” since it implies that the imitator must perform a mental spatial transformation of what he or she sees in order to produce the correct form. To this strategy, we add the caveat that learners must monitor the handedness of the signer and compare it to their own; if hand dominance is discordant, learners may correctly deploy the mirroring strategy.

In addition to these three strategies, yet another imitation strategy is available to learners. Learners may reproduce what they see from their own perspective. This is a *visual matching* strategy because the child's imitative movements match the appearance of what she sees. Let us imagine a child who has adopted this strategy and who is facing the signer. The child sees a sign which originates at a point distal from the signer and which moves toward the signer's own body. The

child could interpret the signer's movement in an absolute sense. She could then reproduce the sign as beginning relatively proximal to her own body and ending at a point distal from her body. Similarly, she could imitate signs exhibiting outward-facing palm orientations (as in BATHROOM, Figure 3) with her palm facing inward toward her own body, thus reversing the palm orientation parameter. Thus, the visual matching strategy would lead to movement and palm orientation errors on signs exhibiting inward–outward movements and palm orientations. Note that this strategy yields predictions about inward–outward movements and palm orientations, but not about hand selection.

To summarize, it appears that children learning sign have at least four possibilities for imitating signs during acquisition:

1. Anatomical strategy: Activate the same muscles as the model, regardless of the hand dominance of the signer.
2. Mirroring strategy: Produce a mirror image of what the signer does.
3. Visual matching strategy: Reproduce the sign as it appears from the learner's perspective.
4. Reversing strategy: Perform a mental spatial transformation on the observed sign and reproduce what the signer does after checking for differences in hand dominance.







Table 1 summarizes each imitation strategy, the conditions under which each strategy will fail, the types of lexical signs that could be susceptible to error when employing each strategy, and the error types that are predicted.

Which strategy or strategies do sign learners adopt, and how would we know? We predict that the difficulties posed by the interaction of sign type (one-handed, two-handed symmetrical, and two-handed asymmetrical), natural variation in handedness, and the four imitation strategies available to learners will lead some sign learners to make specific types of errors, namely *hand-switches*, *lateral and inward–outward movement reversal errors*, and *inward–outward palm orientation reversal errors*, depending on the type of strategy or strategies adopted. We first turn to evidence from published studies on gesture imitation and sign acquisition by typical and atypical hearing and deaf children, as well as by typical adult learners. We then present two new studies of gesture imitation by non-signers, sign learners, and fluent signers to show how exposure to a sign language changes how adults approach imitation. Throughout we demonstrate that hearing and deaf, typical and atypical, children and adults produce errors that reveal the specific difficulties presented by learning a visual language.

STUDIES OF GESTURE IMITATION AND SIGN ACQUISITION BY TYPICAL HEARING AND DEAF CHILDREN






Children begin imitating the actions and gestures of the people around them early in development, for example producing early communicative pointing gestures and conventional gestures such as the “wave bye-bye” gesture by 12 months (Bates, 1979; Carpenter et al., 1998). Studies of the ways that typical infants imitate others suggest that they may shift from an

TABLE 1 | Four imitation strategies in sign learning and predicted error types.

Strategy	Under what circumstances will this strategy lead to errors in sign formation?	Predicted errors	Example signs
Anatomical: Activate the same muscles as the model	When imitating one-handed and two-handed asymmetrical signs if handedness of signer and learner is discordant	Hand switches on one-handed signs such as BATHROOM and two-handed asymmetrical signs such as TOMATO	Sign: BATHROOM (produced by a left-handed signer)
			 Produced (by a right-handed signer) as: 
Mirroring: Produce a mirror image of what the signer does	When imitating signs exhibiting lateral movements if handedness of signer and learner is concordant; for one-handed signs if handedness of signer and learner is concordant	Lateral movement reversals (on signs such as BLACK, SUMMER, BECAUSE, FARM, UGLY, DRY, WE, COMMITTEE, CONGRESS, BOARD, SENATE, ATLANTA, TORONTO, and POLAND); hand switches on one-handed signs such as BATHROOM	Sign: BLACK
			 Produced as: 
			Sign: BATHROOM
			 Produced (by a right-handed signer) as: 

(Continued)

TABLE 1 | Continued

Strategy	Under what circumstances will this strategy lead to errors in sign formation?	Predicted errors	Example signs
Visual matching: Reproduce the sign as it appears from the learner's perspective	When imitating signs exhibiting inward–outward movements and palm orientations; when imitating signs exhibiting lateral movements if handedness of signer and learner is concordant	Inward–outward palm orientation reversals on signs such as TUESDAY and BATHROOM; inward–outward movement reversals on signs such as WANT; lateral movement reversals	Sign: BATHROOM  Produced as: 
Reversing: Reproduce what the signer does after performing a mental spatial transformation and checking for handedness ⁵	Never	None	Sign: BATHROOM  Produced as:  or, for left-handed signers: 

initial visual matching strategy to a mirroring strategy in the second year of life. Evidence for this hypothesis comes from studies on infants' ability to perform role reversal imitation (Tomasello, 1999; Carpenter et al., 2005), that is, performing an action toward another person in the same way that the action

⁵Note that a right-dominant signer can correctly imitate a left-handed model (or a left-dominant signer, a right-handed model) using the mirroring strategy. Under this scenario, a signer would monitor the hand dominance of the model and then employ mirroring if handedness is discordant.

was performed on the child. Two kinds of role reversal imitation have been identified: self-self role reversal, in which the child performs an action on his own body in imitation of an action that the adult performed on her own body (e.g., the infant pats his own head after the adult pats her own head), and other-other role reversal, in which the child performs an action on the adult's body in imitation of an action that the adult performed on the child's body (e.g., the infant pats the adult's head after the adult pats the infant's head). Carpenter et al. (2005) found that

50% of their sample of typical 12-month-old infants and 90% of their sample of typical 18-month-olds performed self-self role reversals, suggesting that this ability develops and strengthens during the second year. The ability to perform such reversals could be key for the development of the reversing or mirroring strategies of imitation, as children imitate not just what they *see* but what others *do*. By the time typical children are preschool age, they are able to imitate the actions of others with high fidelity (Ohta, 1987), and no longer appear to engage the visual matching strategy of imitation.

How does the child's ability to imitate action contribute to the acquisition of signs? Movement errors have frequently been reported for young, typical deaf children acquiring sign (Siedlecki and Bonvillian, 1993; Marentette and Mayberry, 2000; Meier, 2006; Morgan et al., 2007). A problem in interpreting this literature, which is largely based on the observation of naturalistic data, is that it can be difficult to separate errors that arise due to young children's immature motor control from errors that arise due to the perceptual challenges that are the subject of this paper. Crucially, there are very few reports of the development of palm orientation. Palm orientation is the parameter that could shed the most light on the issues raised in this paper because data on inward-outward palm orientations can tell us if children have adopted a visual matching strategy (in which they are likely to reverse inward-outward palm orientations) or have acquired either a reversing or a mirroring strategy (both of which would result in the correct imitation of palm orientation).

To address this question, Shield and Meier (2012) examined 659 tokens in the database of children's early sign productions of four typical deaf children between 9 and 17 months of age (on which Cheek et al., 2001, had based a previous report). This examination revealed 14 tokens (6 inward substitutions and 8 outward substitutions) of reversed inward-outward palm orientation in a database of 659 signs produced by typical deaf children in the first year and a half of life. Thus, it appears that very young typical deaf children do sometimes reverse the palm orientation parameter in a way that appears consistent with their use of the visual matching imitation strategy. However, it is unclear if they do so systematically. We do not yet have a systematic, longitudinal examination of children's acquisition of those sign types that are directly relevant to the hypotheses presented above.

One feature of the way in which infants are socialized to language may contribute to their reconciliation of the different appearances that an individual sign has when viewed from different perspectives. Infants are sometimes seated opposite their parent—say, when they are in a high chair being fed. But infants may also be seated on the parent's lap; in this instance their perspective on the world is closely aligned with that of the parent. Several studies of child-directed signing by Deaf caregivers have shed light on these interactions. Maestas y Moores (1980) studied how American Deaf parents interacted with their infants ($n = 7$); the infants ranged in age from less than a month to 16 months. She found that Deaf parents commonly signed in front of the infant while the infant was seated on the parent's lap such that the viewpoint of parent and infant were shared. Parents also commonly signed on the infants'

bodies, molded their hand configurations, and guided their hand movements, thus providing kinesthetic as well as visual feedback to their children. Similar results have been found for Deaf British mothers who use British Sign Language (Woll et al., 1988).

In a later study, Holzrichter and Meier (2000) reported that four Deaf mothers of deaf children between 8 and 12 months of age displaced signs with a place of articulation on the face onto their children's bodies about 18% of the time (21 of 116 tokens); these instances occurred when there was no eye contact between parent and child during or before the articulation of the sign, such as when the child was sitting on the mother's lap facing away from her. Pizer et al. (2011) describe an interesting example of such an interaction between a Deaf mother and her 18-month-old deaf child. The child was seated on her mother's lap; the mother labeled the colors of the blocks that were on the floor in front of them. The mother produced the ASL signs GREEN, BLUE, and YELLOW in the neutral space in front of the two of them. She then produced the sign ORANGE on her child's mouth rather than on her own body, as normal signing would dictate. Why did the mother do this in this instance? If she had articulated the sign in contact with her own mouth, the sign would not have been visible to the child (because the mother was behind her daughter). When, in these instances, the mother signed GREEN, BLUE, and YELLOW in front of the child and ORANGE on the child's mouth, she enabled her child to witness these signs from the signer's own perspective, rather than from the more typical addressee perspective.

STUDIES OF GESTURE IMITATION AND SIGN ACQUISITION BY HEARING AND DEAF CHILDREN WITH ASD

We also find indications of the challenges presented by learning sign in studies of atypical learners. Children with ASD, both hearing and deaf, show distinctive patterns in imitation.

Hearing Children With ASD

Though language impairment is not considered a core feature of ASD (American Psychiatric Association, 2013), many children with ASD exhibit abnormal language in both speech and sign. A significant minority of children with ASD are considered minimally-verbal (Tager-Flusberg and Kasari, 2013), with expressive vocabularies under 50 words. Manual signs have long been used as an alternative communication strategy for such children, with varying degrees of success (Carr, 1979; Bonvillian et al., 1981, for reviews). In general, minimally-verbal hearing children with ASD are not exposed to, and do not learn, a fully-fledged sign language such as ASL with its syntax and morphology, but instead see a restricted set of lexical signs that are used to communicate basic wants and needs, akin to Baby Signs (Acredolo and Goodwyn, 2002). The published reports on hearing children with ASD who are exposed to signs are unfortunately not useful for the purpose of testing our hypotheses, although Bonvillian et al. (2001) speculate that an unexpectedly high preference for left-handed signing in their

subjects may be attributable to mirroring. We now turn to the literature on gesture imitation by children with ASD.

Various studies have observed that hearing children with ASD do more poorly in general on gesture imitation tasks than typical children, and numerous hypotheses have been advanced to account for these deficits. Edwards (2014) recently performed a meta-analysis of 53 studies on imitation in ASD. She found that individuals with ASD performed on average about 0.8 standard deviations below non-ASD individuals on the imitation tasks contained in the studies, despite important differences between the individual studies depending on the nature of the task and the characteristics of the subject samples. In the section that follows we do not claim to account for all children with ASD, but rather focus on a subset of studies that describe a unique pattern that thus far has only been documented in the imitative behavior of children with ASD.

At least four studies have shown that children with ASD, unlike typical children, produce gesture imitations suggestive of the visual matching imitation strategy. Ohta (1987) was the first to report such errors (which he called “partial imitations”): three of 16 children with ASD between the ages of 6;3 and 14;4 (mean age 10;2) imitated a “wave” gesture (in which the experimenter’s open palm was oriented toward the child) with their palms facing inward toward themselves, consistent with the visual matching imitation strategy. Crucially, no member of an age- and IQ-matched control group or of a second control group of 189 typical preschoolers ages 3–6 imitated the wave gesture in this way, suggesting that this imitation strategy does not occur in typical development beyond a very early age.

Other studies have replicated this striking finding. Smith (1998) found that hearing children with ASD made significantly more 180-degree reversal errors (e.g., palm toward the viewer rather than away from him) than age-matched language-impaired and typically developing children when imitating ASL handshapes and bimanual gestures. Whiten and Brown (1998: 270–271) also found that hearing children with ASD made similar gesture imitation errors, highlighting

responses in which the imitating subject creates an *action which to him will look similar to what he saw when he watched the demonstrator, instead of what the demonstrator would see*. He fails to translate appropriately, or “invert” the action to his own perspective as actor. An example is “peekaboo,” performed by the demonstrator with palms toward her own face, and sometimes inaccurately imitated such that the palms are oriented away from the imitator’s face (i.e., the actor sees the backs of the hands both when the demonstrator performs the act, and when he himself attempts it) (*emphasis ours*).

Adding to these findings, Hobson and Lee (1999) provide a crucial link between the reversal errors in gesture imitation and the role reversal skills described by Tomasello (1999). They found that adolescents with ASD were significantly less likely to imitate a self-oriented action (wiping their own brow with a toy frog after an adult did so) than were age- and language-matched intellectually-disabled children: only five of 16 children with ASD performed the self-oriented action while 14 of 16 of the control children did so. This finding suggests that it is indeed this early

development of role reversal skills that enables typical children to transcend the visual matching strategy. That visual matching strategy has now surfaced in multiple studies of how children with ASD imitate gestures.

Deaf Children With ASD

More recently, Shield and colleagues have published a number of studies describing the acquisition of ASL by deaf children with ASD who have Deaf parents (Shield and Meier, 2012; Shield, 2014; Shield et al., 2015, 2016, 2017a,b; Bhat et al., 2016). The first report (Shield and Meier, 2012) described the formational errors produced by five native-signing children with ASD (four deaf children and one hearing child of Deaf adults) ranging in age from 4;6 to 7;5. These children were compared to a control group of 12 typical native-signing deaf children between the ages of 3;7 and 6;9. The data came from spontaneous signing produced under naturalistic conditions and from a fingerspelling task (in which children were asked to spell English written words with their hands). Despite lifelong exposure to ASL, three of the children with ASD (ages 5;8, 6;6, and 7;5) reversed the palm orientation of 72 of 179 (40.2%) fingerspelled letters such that the children’s palm faced toward their own body rather than outward. None of the 12 typical deaf children produced any such palm orientation reversals. These reversals are consistent with the visual matching strategy of imitation and are nearly identical to the errors produced by hearing, non-signing children with ASD in the previously-discussed studies of gesture imitation (Ohta, 1987; Smith, 1998; Whiten and Brown, 1998). The three children with ASD who made such errors had lower parent-reported language scores ($M = 36.67$, $SD = 13.61$, range 26–52) on the Language Proficiency Profile-2 (LPP-2; Bebko et al., 2003) than those children who did not make such errors, including the 12 typical deaf children ($M = 90.25$, $SD = 17.07$, range 59–112) or the child with ASD who did not make any palm reversals ($=90$). This difference was significant [$t_{(14)} = 5.23$, $p < 0.001$], suggesting that children with lower receptive and expressive language skills may be more prone to making such errors.

If the palm orientation reversals exhibited by native-signing children with ASD are the result of the visual matching imitation strategy, then how do such children perform on gesture imitation tasks? Two studies have shown that even deaf children who are exposed natively to a sign language nonetheless show difficulties with gesture imitation. In his unpublished dissertation, Shield (2010) asked 12 typical deaf children and 17 deaf children with ASD to imitate nonsense signs similar to ASL signs. He divided up the target stimuli into test items (hypothesized to require a reversing strategy in order to be imitated correctly, i.e., with lateral path movements) and control items (which do not require a reversing strategy in order to be imitated correctly, i.e., with up-down path movements). The children with ASD made significantly more imitation errors than typical controls overall, as well as significantly more errors on test items than control items, suggesting that gestures that require a reversing imitation strategy can be particularly difficult for such learners. The children with ASD also had significantly lower language scores on the LPP-2 ($M = 66.25$, $SD = 31.49$) than the typical children ($M = 90.25$, $SD = 17.07$), again indicating a relationship between these errors and overall language abilities. Children with

ASD made significantly more errors on inward–outward palm orientations than on any of the other item types or parameters, which may be a sign of the visual matching imitation strategy. Thus, the observation of these palm orientation reversals in gesture imitation by deaf, signing children provides a link between the reversed signs observed by Shield and Meier (2012) in spontaneous and elicited production of ASL and the reversed gestures observed in hearing children with ASD by Ohta (1987), Smith (1998) and Whiten and Brown (1998). All of the errors indicate that some children with ASD use a visual matching strategy in imitation far beyond the age that typical children stop doing so.

Shield (2010) also examined whether right-handed children switched hands during the task as a way of avoiding the reversing strategy when imitating the right-handed investigator (thereby using the mirroring strategy instead). Both typical and ASD children switched hands significantly more often on test items than control items, suggesting that both groups preferred to avoid the reversing strategy for gestures that were more difficult to imitate. Moreover, younger children switched hands more often than older children, which implies that exposure to and practice with imitation of gestures renders these processes easier over time.

More recently, Shield et al. (2017b) examined the ability of 14 deaf children with ASD between 5 and 14 years old ($M = 9.5$) and 16 age- and IQ-matched typical deaf children to imitate a series of 24 one-handed gestures exhibiting inward–outward movements and palm orientations and up–down movements and palm orientations. They found that children with ASD made significantly more palm orientation errors than typical children (though movement direction errors were largely absent in both groups). Both groups were also inconsistent in the hand that they used to imitate the gestures, possibly to avoid the reversing strategy: on average children with ASD switched hands in 5.67 of 24 trials (23.6%), while typical children switched hands in 3.26 of 24 trials (13.6%). However, note that 10 of 16 typical deaf children and 7 of 14 deaf children with ASD were consistent in using the same hand to imitate all of the trials; these children never switched hands.

Taken together, these studies lead us to think that the imitation of certain types of signs and gestures is particularly difficult for hearing and deaf children with ASD. The inward–outward palm orientation reversal errors identified in studies of gesture imitation by hearing and deaf children with ASD (Ohta, 1987; Smith, 1998; Whiten and Brown, 1998; Shield, 2010; Shield et al., 2017b) and in the sign language of some hearing and deaf children with ASD (Shield and Meier, 2012) suggest that some children with ASD employ the visual matching strategy in gesture imitation, and that this approach to imitation can then influence how children produce signs on their own. Typical children do not appear to employ this strategy once they have mastered role reversal during the very earliest stages of language development. Both typical children and children with ASD switch hands when imitating gestures hypothesized to require the reversing strategy in order to be imitated correctly, thus resorting to the less-difficult mirroring strategy.

STUDIES OF GESTURE IMITATION AND SIGN ACQUISITION BY TYPICAL ADULTS

In this section, we add to the evidence from studies of children, by reviewing several studies of how typical adults learn signs and imitate gestures. We ask if adults who are learning a sign language exhibit patterns like those described for children, and we ask how adults who have no exposure to sign imitate gestures.

Sign Learning

Rosen (2004) studied 21 adult beginning learners of ASL in a 15-week course and described the types of errors they made in articulating signs. He predicted error types based on perceptual and articulatory factors; here we discuss only the former. He noted that perceptual errors would be rooted in “the physical stance from which the learner views the input source such as the teacher” and would occur “when signers either mirror or make parallel their signs with those of the teacher” (p. 38). Such perceptual errors could then lead to a situation wherein “signers may reverse the handshape, location of contacts, direction of movements, and the orientation of palms within lexical signs as compared to their teacher” (p. 38). As he predicted, Rosen found that adult learners of sign made location, movement, and palm orientation errors based on what he called “mirrorization” and “parallelization.” In our terminology, “mirrorization” errors reflect either the mirroring or anatomical strategy and “parallelization” errors reflect the visual matching strategy. Mirroring errors included reversals of lateral movements; anatomical errors were evidenced by hand switches from dominant to non-dominant. Visual matching errors included palm orientation reversal errors such as the ASL sign *DOOR* produced with palms facing inward rather than outward. Thus, Rosen found that adult learners of sign struggled with particular types of signs and utilized, in our terms, the mirroring, anatomical, and visual matching strategies to produce them. Unfortunately, he included no quantitative analyses so we do not know how frequently the beginning learners made such errors. Nonetheless, the documentation of these error types in the literature is helpful insofar as it suggests that some signs are more difficult to learn than others, and that typical adults employ several of the imitation strategies we describe in this paper.

Gesture Imitation

We again look to studies of gesture imitation to verify if the errors observed in sign production could be the result of imitation processes. Shield (2010) asked 24 hearing, right-handed undergraduate students who were naive to sign to imitate 48 manual gestures, half of which were extant ASL signs and half of which were nonsense gestures created by modifying the ASL signs. By hypothesis, half of the gestures required a reversing strategy in order to be imitated correctly (i.e., lateral and inward–outward path movements and inward–outward palm orientations) and half did not (i.e., up–down path movements and palm orientations).

The undergraduates made significantly more errors when imitating ASL signs and nonsense gestures hypothesized to require the reversing strategy in imitation (e.g., exhibiting a

lateral path movement) than on control items. Signs involving a lateral movement were particularly vulnerable to error: 23 of 44 tokens (52.3%) of the sign **BLACK** (which moves from the contralateral side of the forehead to the ipsilateral side; see **Table 1**) contained a movement reversal error, while 16 of 44 tokens (36.4%) of the sign **FLOWER**, which also entails a lateral path movement across the face, contained a movement reversal error. Two of the subjects imitated all gestures with their left hand (despite being right-handed), thus employing the mirroring strategy and avoiding the reversing strategy. Unlike children with ASD, however, the undergraduates had no difficulty with inward-outward movements or palm orientations and did not appear to use the visual matching strategy.

Thus, this study suggests that typical adults tend to engage a mirroring strategy when imitating novel gestures, which is successful except in the case of lateral path movements when handedness is shared between model and subject. In such cases, typical adults made lateral movement reversal errors or switched hands in order to mirror the gesture correctly.

DOES SIGN LANGUAGE EXPOSURE CHANGE HOW LEARNERS IMITATE?

We have shown that the mirroring and visual matching strategies both lead to specific kinds of imitation errors; furthermore it appears that typical and atypical children as well as typical adults produce errors consistent with these strategies in their signing and gesture imitation. We now present two new studies to further examine our hypotheses. We ask if sign language exposure can change how learners imitate gestures. Specifically, we hypothesize that sign language exposure could shift typical learners from a mirroring strategy to a reversing strategy due to practice with reversing.

Study 1: Mirroring and Signer Experience Methods

To test the hypothesis that sign exposure may enable typical adult learners to shift from a mirroring strategy to a reversing strategy, we recruited non-signers, sign learners (intermediate ASL students), and fluent signers for a study of gesture imitation.

Stimuli

We created 48 gesture stimuli based on four palm orientations (up, down, in, out), six movements (inward toward the body, outward from the body, up, down, ipsilateral→contralateral, contralateral→ipsilateral), and two handshapes (the 1- and 5-handshapes); see **Table 2**. Each palm orientation type was combined with each movement type to create 24 base gestures; each of these gestures was then filmed twice, once with a 1-handshape (with the index finger extended and all other fingers retracted) and again with the 5-handshape (with all fingers extended). Each videotaped stimulus lasted 1.5 s. None of the gestures were extant ASL signs; thus, they were meaningless for signers and non-signers alike.

We hypothesized that all subjects would be able to imitate gestures with vertical and horizontal movements since these can be imitated using the mirroring strategy. However, we predicted

that subjects with exposure to ASL would imitate gestures with lateral movements more accurately than non-signers, since these must be imitated using the reversing strategy. We did not predict that any of the groups would have difficulty with the four palm orientations, since these can also be imitated using a mirroring strategy. In order to ensure that all participants would have the opportunity to engage the reversing strategy, we verified the handedness of each participant and then used either a right- or left-handed version of the stimuli, such that every participant imitated a model with concordant handedness. The left-handed version of the stimuli was made by flipping the right-handed stimuli horizontally; thus, the stimuli presented to left- and right-handed participants were identical in every aspect, save for the apparent handedness of the model.

Procedure

Participants stood in front of a 17" MacBook laptop computer, which was placed approximately at eye level three feet away. Participants were instructed to reproduce each gesture as accurately as possible. Each participant viewed each of the 48 gesture stimuli in one of two pre-established random orders; no stimuli were repeated. A 3-s pause followed each gesture stimulus during which participants were asked to imitate the gesture observed.

Participants

We recruited three groups of participants: (1) non-signing undergraduate students at Boston University who had never had any exposure to sign language ($N = 34$; all right-dominant, 19 females), (2) sign learners, students who were then enrolled in the fourth or fifth semester of an ASL course ($N = 25$; 23 right-dominant, 22 females), and (3) fluent signers, either professional sign language interpreters or Deaf adults ($N = 18$; all right-dominant, 12 females)⁶.

Coding



















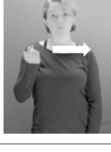




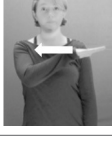
Each trial was coded blindly by a Deaf native signer for movement direction and palm orientation values so that the coder did not know what the stimulus gesture had been. There were two values per stimulus, a movement value and a palm orientation value. A second coder, a fluent signer, then matched the coded trials to the target movement and palm orientation values and re-coded each trial as correct or incorrect. Any movement or palm orientation value other than the target was considered an error. In order to assess intercoder reliability, a third coder (also a fluent signer) re-coded 20% of the trials. There were 10 disagreements out of 288 re-coded trials; Cohen's κ was 0.97 for palm orientation (6 disagreements out of 288 trials) and 0.98 for movement (4 disagreements out of 288 trials), indicating very high levels of agreement.

Statistical analysis

We fit a generalized linear mixed-effects model using error frequency as the dependent variable. The independent variables

⁶Although we did not collect data on the ages of participants, it is worth noting that the fluent signers were working adults or graduate students, while the other two groups were undergraduate students.

TABLE 2 | Gesture stimulus types by palm orientation and direction of movement.

MOVEMENT			Palm orientation			
			Horizontal		Vertical	
			Out	In	Up	Down
	Vertical	Up				
		Down				
	Horizontal	In				
		Out				
	Lateral	Ipsi-Contra				
		Contra-Ipsi				

Every gesture type above was shown to participants twice, once with a 1-handshape and once with a 5-handshape.

^aExample of a stimulus with a 1-handshape (index finger extended).

^bExample of a stimulus with a 5-handshape (all five fingers extended).

were experience (non-signer, sign learner, or fluent signer) and gesture type (vertical, horizontal, or lateral movements; up–down or in-out palm orientation). The mixed effects were necessary to model the repeated measures design of the gesture type variable.

Results

Non-signers erred on 6.85% of the 48 gestures imitated ($M = 6.56$ errors, $SD = 4.62$), sign learners erred on 2.63% of gestures imitated ($M = 2.52$ errors, $SD = 2.29$), and fluent signers erred on 1.39% of gestures imitated ($M = 1.33$ errors, $SD = 1.88$). Experience was a significant predictor of performance, $X^2_{(2)} = 36.03$, $p < 0.0001$. *Post-hoc* Tukey comparisons found that non-signers produced significantly more errors than either sign learners ($z = 4.30$, $p < 0.001$) or fluent signers, ($z = 5.59$, $p < 0.001$). The difference between the sign learners and the fluent signers was not quite significant ($z = 2.08$, $p = 0.09$).

Movement items

Non-signers produced a significantly higher error rate (24.3%; $M = 3.88$ errors, $SD = 3.41$) than either sign learners (14.3%, $M = 2.28$ errors, $SD = 2.3$) or fluent signers (6.3%, $M = 1.0$ errors, $SD = 1.68$) on lateral (ipsilateral–contralateral or vice versa) movements [$X^2_{(2)} = 17.23$, $p < 0.001$], see **Figure 4**. Non-signers also produced a significantly higher error rate (2.81%, $M = 0.45$ errors, $SD = 0.88$) than either sign learners (0%) or fluent signers (0%) on inward–outward movements [$X^2_{(2)} = 10.20$, $p < 0.01$]. There were no group differences in error rates on up–down movements; the non-signers produced two total errors on this parameter, while the sign learners and fluent signers did not produce any errors on this parameter.

Palm orientation items

Non-signers made more palm orientation errors than either sign learners or fluent signers for both up–down [$X^2_{(2)} = 11.80$,

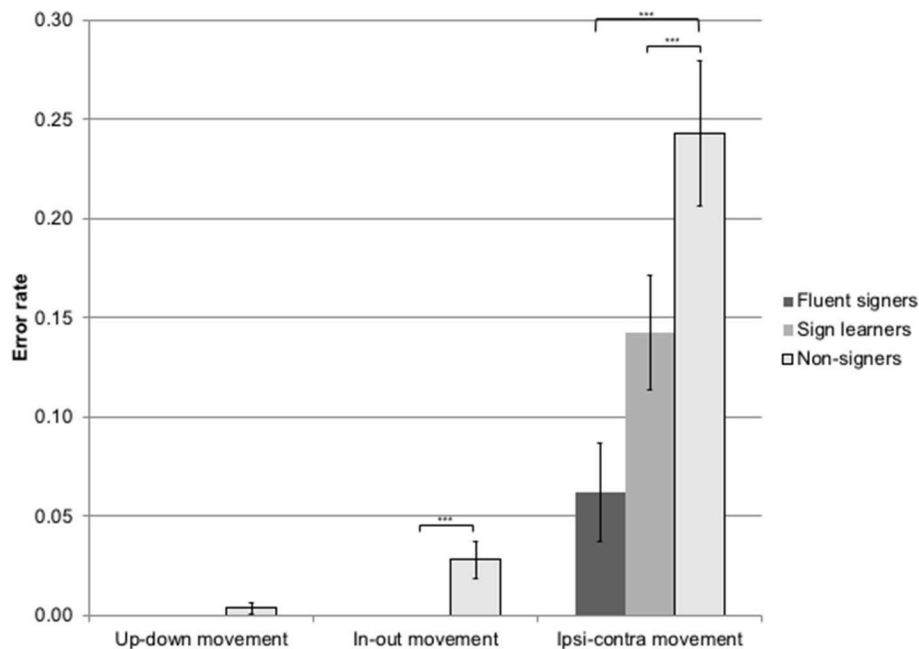


FIGURE 4 | Error rate by movement type for non-signers, sign learners, and fluent signers. *** $p < 0.001$.

$p < 0.01$] and in-out [$X^2_{(2)} = 20.43, p < 0.0001$] palm orientation types. Non-signers produced an error rate of 5.04% on up-down palm orientations ($M = 1.2$ errors, $SD = 1.1$), compared to 0.83% ($M = 0.2$ errors, $SD = 0.1$) for sign learners and 1.16% ($M = 0.28$ errors, $SD = 0.57$) for fluent signers. On in-out palm orientations, non-signers produced an error rate of 4.04% ($M = 0.97$ errors, $SD = 1.58$) compared to 0.17% ($M = 0.04$ errors, $SD = 0.2$) for sign learners and 0.23% ($M = 0.06$ errors, $SD = 0.24$) for fluent signers; see **Figure 5**.

Discussion

We predicted that subjects would make more imitation errors on gestures involving lateral movements across the body than on gestures involving vertical or horizontal movements due to a bias toward a mirroring strategy rather than a reversing strategy. Our prediction was borne out: lateral movements were significantly more susceptible to error than other movement types. We further predicted that imitation performance would interact with exposure to sign language, with fluent signers making the fewest number of errors, followed by sign learners, and finally by non-signers (though note that we did not detect statistical differences between the sign learners and the fluent signers). This prediction was also borne out both for the movement and palm orientation gesture types. In particular, non-signers produced a significantly higher rate of reversals on lateral movement gestures (24%) than sign learners (14%) or fluent signers (6%). Non-signers also produced more errors on horizontal (in-out) movements than either sign learners or fluent signers. Importantly, neither signers nor non-signers made errors on the control condition of imitating up-down (vertical) movements.

Non-signers also produced more errors on both kinds of palm orientations than either sign learners or fluent signers. We did not predict these error types; one plausible explanation for their occurrence is that non-signers may have been paying particular attention to the more perceptually salient movements and were paying insufficient attention to palm orientation. Subjects with sign exposure know to pay attention to both movement and palm orientation, since both have linguistic value in sign.

Study 1 showed that certain types of gesture found in signed languages are more difficult to imitate, especially for non-signers who tend to employ the mirroring strategy, leading to lateral movement errors. However, the reversing strategy is only necessary when imitating lateral movements produced by people with the same hand dominance, i.e., right-handers imitating right-handers or left-handers imitating left-handers. Would right-handed non-signers still make more errors on lateral movements if they were imitating a left-handed model, and thus could use a mirroring imitation strategy? In order to test this specific hypothesis, we designed an additional study to examine the role that handedness plays in perspective-taking.

Study 2: Mirroring and Discordant Handedness

If the difficulty of the reversing strategy is truly at issue in the imitation of lateral movement gestures, then right-handed subjects should only have difficulty imitating other right-handers. Thus, we predicted that right-handed subjects would not have a problem imitating lateral gestures produced by a left-handed model, since such movements can be imitated with a mirroring strategy rather than a reversing strategy.

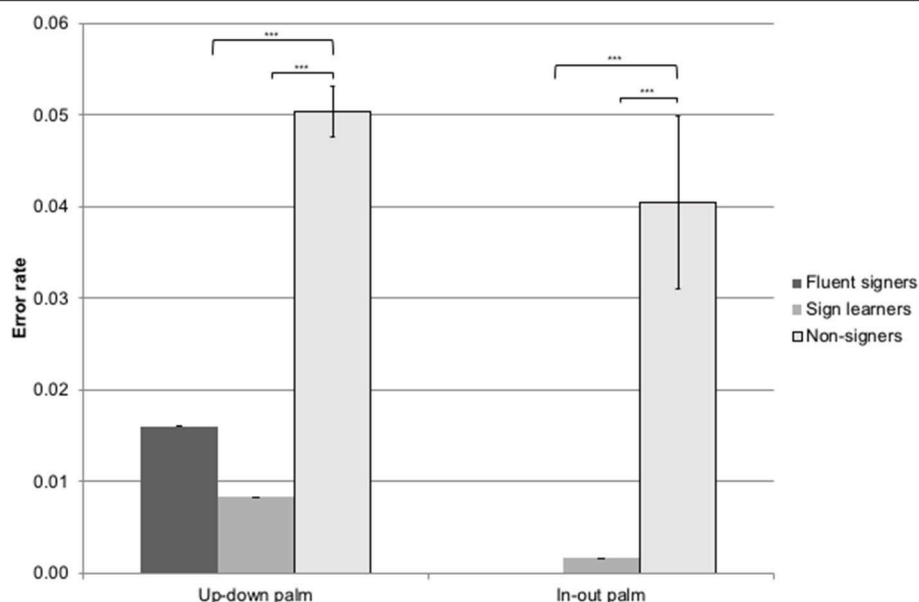


FIGURE 5 | Error rate by palm orientation type for non-signers, sign learners, and fluent signers. *** $p < 0.001$.

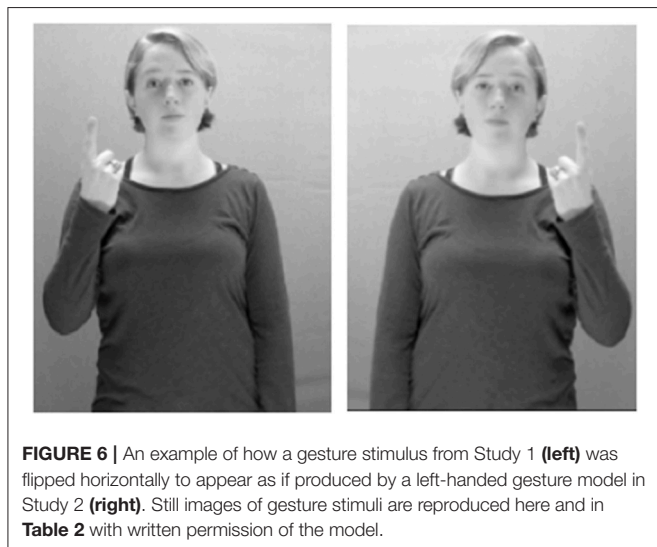


FIGURE 6 | An example of how a gesture stimulus from Study 1 (left) was flipped horizontally to appear as if produced by a left-handed gesture model in Study 2 (right). Still images of gesture stimuli are reproduced here and in Table 2 with written permission of the model.

Methods

To test the hypothesis that discordant handedness allows imitators to avoid the reversing strategy on difficult lateral movements, we modified the stimuli used in Study 1.

Stimuli

The 48 gesture stimuli used in Study 1 were flipped horizontally such that it now appeared that the right-handed gesture model was producing the gestures with her left hand; see Figure 6. We predicted that right-handed non-signers would not make lateral movement errors in this condition, since they should be able to use mirroring to correctly imitate.

Subjects

For Study 2 we recruited 67 non-signing undergraduate students; 34 right-handed non-signers (19 women) were assigned at random to the flipped condition, and 33 right-handed non-signers (27 women) were assigned at random to the same non-flipped condition as in Study 1.

Results

Results for Study 2 are shown in Figure 7. In the flipped condition, participants made 10 errors on vertical movements out of 544 trials (1.8%), while in the non-flipped condition, participants made 2 errors on vertical movements out of 528 trials (0.4%). The difference between conditions for vertical movements was marginally significant (Fisher's Exact Test, $p = 0.05$). On horizontal movements, participants in the flipped condition made 6 errors (1.1% of 544 trials); participants in the non-flipped condition also made 6 errors on horizontal movements (1.1% of 528 trials). There was no difference between the two conditions for horizontal movements (Fisher's Exact Test, $p = 1.0$, ns). On lateral movements, participants in the flipped condition made just 5 errors (0.9% of 544 trials), but 109 errors in the non-flipped condition (20.6% of 528 trials). A two-sample Cramer Von-Mises test found that error rate on lateral movements was significantly lower ($p < 0.001$) in the flipped condition than in the non-flipped condition.

No differences were detected between the error rates for palm orientations. Participants produced errors on 1.64% of up-down palm orientations in the non-flipped condition and 2.82% in the flipped condition (Fisher's Exact Test, ns), and 3.41% of in-out palm orientations in the non-flipped condition and 1.47% in the flipped condition (Fisher's Exact Test, ns).

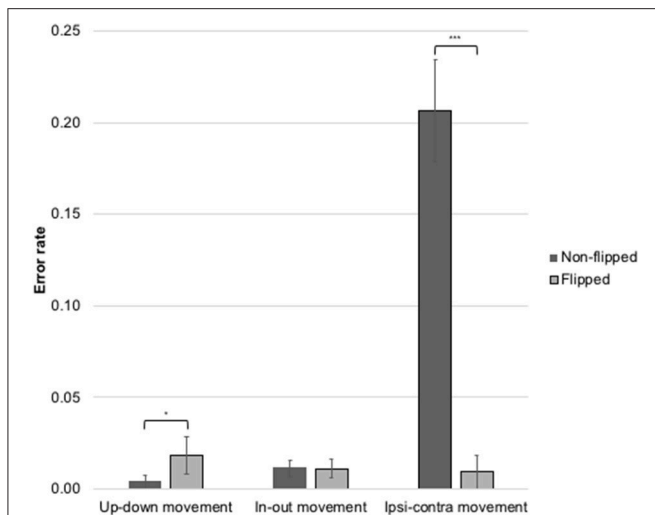


FIGURE 7 | Error rates on movement types in the flipped and non-flipped conditions. Subjects were sign-naïve undergraduates. *** $p < 0.001$, * $p = 0.05$.

Discussion

Study 2 showed that handedness interacts with imitation in specific and predictable ways. First, subjects in the non-flipped condition exhibited a high error rate (20.6%) when imitating lateral movements, replicating the results of Study 1 and confirming that imitating these gestures is difficult. Second, subjects in the flipped condition (who thus appeared to be imitating a left-handed model) made significantly fewer errors (0.9% error rate). We thus demonstrate that gestures exhibiting lateral path movements can be successfully imitated by right-handed subjects when the model being imitated performs the movements with her left hand, thereby enabling a *mirroring* strategy rather than a *reversing* strategy. Thus, we find strong evidence that gesture imitation strategies are influenced by handedness and that lateral movements are easier for non-signers to imitate when handedness is discordant, in line with our predictions. We find no difference in the flipped and non-flipped conditions for palm orientation, in accordance with our prediction that palm orientation would not be affected by the handedness of the model.

The two new gesture imitation studies described here support three hypotheses about the difficulties involved in learning a sign language. First, sign language exposure changes how adults approach imitation, shifting them from a mirroring strategy to a more difficult reversing strategy. Second, lateral movements across the body are more difficult to imitate than either horizontal (inward–outward) or vertical (up–down) movements, since they require a reversing strategy in order to be successfully imitated, provided that the handedness of the imitator and the model is concordant. Since left dominance is relatively rare, concordant handedness is likely to be true of the large majority of sign learning encounters. Third, we demonstrate the role that handedness plays in the imitation of lateral movements, as right-handed non-signers were significantly better at imitating lateral movements when imitating an apparently left-handed model.

GENERAL DISCUSSION

We have described some of the ways in which language acquisition in the visual-gestural modality poses unique challenges for sign language learners. Our argument can be summarized as follows:

1. Signed languages use space for several purposes. They can use space to talk about space, as in the descriptions of spatial arrays discussed by Emmorey (2002). They can use space to mark grammatical relations, as in verb agreement and anaphora. Finally, they can use space in a fixed way, as in lexical signs. The sign-learning child must learn to distinguish these different constructions and uses of space. Spatial arrays and grammatical uses of spatial anaphora are relatively advanced skills that appear later in development, but the acquisition of lexical signs occurs early. Children must figure out that lexical signs are not specified for right and left (unlike the spatial layout depicted in **Figure 2**) but, instead for the movements of the dominant and non-dominant hands.

2. The sign lexicon is composed of different types of signs. Some are one-handed and some are two-handed; two-handed signs may be symmetrical (with both hands exhibiting the same handshapes and movements) or asymmetrical (with each hand exhibiting a different handshape and movement). Signers vary in hand dominance, thus input to children is varied in terms of how they see one-handed and two-handed asymmetrical signs being produced. Children also view signs from various perspectives, further complicating the input they receive.

3. At least four imitation strategies are available for imitating signs. One strategy is the anatomical imitation strategy, in which subjects activate the same muscles as the model they are imitating, resulting in the switching of the hands from dominant to non-dominant when signer and model do not share handedness. We find evidence that typical adults, as well as typical and atypical children, sometimes use this strategy, particularly when imitating difficult gestures. A second strategy is the mirroring strategy, in which subjects produce a mirror image of the gestures or signs they are imitating. We find evidence that typical adults learning sign and imitating gestures tend to use this strategy, and that this results in lateral movement errors when handedness is shared. A third strategy is the visual matching strategy, in which subjects imitate what they see from their own perspective. This leads to reversals in inward–outward palm orientations and inward–outward movements in gesture imitation and sign production. We find evidence that typical adults learning sign, very young typical children, and older hearing and deaf children with ASD sometimes employ this strategy. Finally, skilled signers employ a reversing strategy, in which they perform a mental spatial transformation in order to reproduce the model's gesture. We find that fluent sign language users and sign language learners are better at imitating gestures using the reversing strategy than are non-signers, who prefer the mirroring strategy. We thus find evidence that sign language exposure changes the way that typical adults imitate gestures.

CONCLUSION AND FUTURE DIRECTIONS

We argue that the visual-gestural modality presents challenges to sign language learners unlike the challenges faced by learners of spoken languages. Learners confront variation in input due to differences in handedness in the population, with no obvious analog in speech. One potential analog in speech is the acoustic variation in phoneme production caused by differently-sized vocal tracts, but it is unclear how comparable these two phenomena are. Furthermore the visual-gestural modality allows for multiple ways to interpret the imitation task, while the vocal-auditory modality generally does not. An exception in speech arises in the imitation of pronouns, such that an imitation of the sentence “Mommy loves you” can retain the modeled pronoun or can replace it with “me”, thereby preserving the reference of the model sentence.

In the future we need further work on the acquisition of the sign lexicon by typical deaf children. In particular we need better documentation of their early sign development with regard to the specific predictions made here, especially with respect to the movement and palm orientation types discussed. It would also be interesting to know if signs hypothesized to be difficult to imitate are acquired relatively late in development. A systematic analysis of the MacArthur-CDI database for ASL signs (Anderson and Reilly, 2002; <http://wordbank.stanford.edu>) could shed light on this problem.

The reports of reversed inward-outward palm orientations in children with ASD, whether hearing children imitating gestures or deaf children producing signs, are a robust indicator that some children with ASD use the visual matching strategy in imitation. However, we still do not have a clear understanding of which children with ASD tend to use this strategy nor how frequently the phenomenon occurs. It may just be a subset of children with ASD who use this strategy rather than being a characteristic strategy of all children with ASD; a crucial question to ask is if those children who employ the visual matching strategy also share a cognitive profile and if other related cognitive characteristics can be identified.

Lastly, we need further work on the gesture development of hearing children in the first 2 years of life. We need systematic

documentation of whether or not typical infants reverse the direction of their palm when producing early gestures such as the “wave bye-bye” gesture, when they produce the gesture in its mature form, and what other cognitive milestones occur contemporaneously. Such work on typical deaf and hearing children will put us in a better position to understand the development of sign and gestures in children with ASD. It will also help clarify how children approach imitation and if emergent imitation strategies can be more clearly linked to sign language development.

ETHICS STATEMENT

All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the Institutional Review Board of Boston University.

AUTHOR CONTRIBUTIONS

AS designed and conducted the original studies reported in this paper, contributed to the development and refining of the hypotheses described in the paper, and was the primary author of the paper. RM contributed to the development and refining of the hypotheses described in the paper, the designing of the original studies reported, and the writing of the paper.

FUNDING

This work was supported by grant 1F32-DC0011219 from NIDCD and Research Enhancement Grant 14-04 from the Autism Science Foundation to AS.

ACKNOWLEDGMENTS

We thank T. Sampson and A. Hough for coding data, A. Hough for modeling gestures, F. Ramont for modeling ASL signs, A. Marks for taking photos of ASL signs, N. Coffin, A. Hensley, and H. Harrison for statistical consulting and analysis, B. Bucci and the Deaf Studies Program at Boston University for recruitment of ASL students, and H. Tager-Flusberg for research support.

REFERENCES

- Acredolo, L. P., and Goodwyn, S. (2002). *Baby Signs: How to Talk With Your Baby Before your Baby Can Talk*. New York, NY: McGraw-Hill Professional.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, Arlington, VA: American Psychiatric Publishing.
- Anderson, D., and Reilly, J. (2002). The macarthur communicative development inventory: normative data for American sign language. *J. Deaf Stud. Deaf Educ.* 7, 83–106. doi: 10.1093/deafed/7.2.83
- Bates, E. (1979). *The Emergence of Symbols: Cognition and Communication in Infancy*. Cambridge, MA: Academic Press.
- Battison, R. (1978). *Lexical Borrowing in American Sign Language*. Silver Spring, MD: Linstok Press.
- Bebko, J. M., Calderon, R., and Treder, R. (2003). The Language Proficiency Profile-2: Assessment of the global communication skills of deaf children across languages and modalities of expression. *J. Deaf Stud. Deaf Educ.* 8, 438–451. doi: 10.1093/deafed/eng034
- Bellugi, U., O’Grady, L., Lillo-Martin, D., O’Grady, M., van Hoek, K., and Corina, D. (1990). “Enhancement of spatial cognition in deaf children,” in *From Gesture to Language in Hearing and Deaf Children*, eds V. Volterra and C. Erting (Washington, DC: Gallaudet University Press), 278–298.
- Bhat, A. N., Srinivasan, S. M., Woxholdt, C., and Shield, A. (2016). *Differences in Praxis Performance and Receptive Language During Fingerspelling Between deaf Children with and Without Autism Spectrum Disorder*. Autism: Advance online publication.
- Bonvillian, J. D., Gershoff, E. T., Seal, B. C., and Richards, H. C. (2001). Hand preferences in sign-learning students with autistic disorder. *Laterality* 6, 261–281. doi: 10.1080/713754414
- Bonvillian, J. D., Nelson, K. E., and Rhyne, J. M. (1981). Sign language and autism. *J. Autism Dev. Disord.* 11, 125–137. doi: 10.1007/BF01531345
- Bonvillian, J. D., Orlansky, M. D., and Garland, J. B. (1982). Handedness patterns in deaf persons. *Brain Cogn.* 1, 141–157. doi: 10.1016/0278-2626(82)90012-4

- Butterworth, G., and Hopkins, B. (1993). Origins of handedness in human infants. *Dev. Med. Child Neurol.* 35, 177–184. doi: 10.1111/j.1469-8749.1993.tb11620.x
- Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63, 1–143. doi: 10.2307/1166214
- Carpenter, M., Tomasello, M., and Striano, T. (2005). Role reversal imitation and language in typically developing infants and children with autism. *Infancy* 8, 253–278. doi: 10.1207/s15327078in0803_4
- Carr, E. G. (1979). Teaching autistic children to use sign language: Some research issues. *J. Autism Dev. Disord.* 9, 345–359. doi: 10.1007/BF01531444
- Cheek, A., Cormier, K., Repp, A., and Meier, R. P. (2001). Prelinguistic gesture predicts mastery and error in the production of early signs. *Language* 77, 292–323. doi: 10.1353/lan.2001.0072
- Conrad, R. (1979). *The Deaf Schoolchild: Language and Cognitive Function*. London: Harper & Row.
- Corballis, M. (1980). Laterality and myth. *Am. Psychol.* 35, 284–295. doi: 10.1037/0003-066X.35.3.284
- Corballis, M. (1992). *The Lopsided Ape*. New York, NY: Oxford University Press.
- Courtin, C., and Melot, A. M. (1998). “Development of theories of mind in deaf children,” in *Psychological Perspectives on Deafness*, eds M. Marschark and M. D. Clark (Mahwah, NJ: Lawrence Erlbaum Associates), 79–102.
- Edwards, L. A. (2014). A meta-analysis of imitation abilities in individuals with autism spectrum disorders: meta-analysis of imitation in ASD. *Autism Res.* 7, 363–380. doi: 10.1002/aur.1379
- Emmorey, K. (2002). “The effects of modality on spatial language: How signers and speakers talk about space,” in *Modality and Structure in Signed and Spoken Languages*, eds R. P. Meier, K. Cormier, and D. Quinto-Pozos (Cambridge: Cambridge University Press), 405–421.
- Emmorey, K., Bosworth, R., and Kraljic, T. (2009). Visual feedback and self-monitoring of sign language. *J. Mem. Lang.* 61, 398–411. doi: 10.1016/j.jml.2009.06.001
- Emmorey, K., Klima, E., and Hickok, G. (1998). Mental rotation within linguistic and non-linguistic domains in users of American Sign Language. *Cognition* 68, 221–246. doi: 10.1016/S0010-0277(98)00054-7
- Emmorey, K., Tversky, B., and Taylor, H. A. (2000). Using space to describe space: perspective in speech, sign, and gesture. *Spat. Cogn. Comput.* 2, 157–180. doi: 10.1023/A:1013118114571
- Franz, E. A., Ford, S., and Werner, S. (2007). Brain and cognitive processes of imitation in bimanual situations: making inferences about mirror neuron systems. *Brain Res.* 1145, 138–149. doi: 10.1016/j.brainres.2007.01.136
- Hamilton, A. F., Brindley, R., and Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 37–44. doi: 10.1016/j.cognition.2009.07.007
- Hobson, R. P., and Lee, A. (1999). Imitation and identification in autism. *J. Child Psychol. Psychiatry* 40, 649–659. doi: 10.1111/1469-7610.00481
- Hoffmann-Dilloway, E. (2017). Feeling your own (or someone else's) face: Writing signs from the expressive viewpoint. *Lang. Comm.* doi: 10.1016/j.langcom.2016.12.011. [Epub ahead of print].
- Holzrichter, A. S., and Meier, R. P. (2000). “Child-directed signing in American Sign Language,” in *Language Acquisition by Eye*, eds C. Chamberlain, J. P. Morford and R. I. Mayberry (Mahwah, NJ: Lawrence Erlbaum Associates), 25–40.
- Klima, E. S., and Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.
- Koski, L. (2003). Modulation of cortical activity during different imitative behaviors. *J. Neurophysiol.* 89, 460–471. doi: 10.1152/jn.00248.2002
- Kuhl, P. K. (1979). Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *J. Acoust. Soc. Am.* 66, 1668–1679. doi: 10.1121/1.383639
- Lillo-Martin, D., Bellugi, U., Struxness, L., and O'Grady, M. (1985). The acquisition of spatially organized syntax. *Papers Rep. Child Lang. Dev.* 24, 70–78.
- MacNeillage, P. (2008). *The Origin of Speech*. Oxford: Oxford University Press.
- Maestas y Moores, J. (1980). Early linguistic environment: interactions of Deaf parents with their infants. *Sign Lang. Stud.* 26, 1–13. doi: 10.1353/sls.1980.0003
- Marentette, P. F., and Mayberry, R. I. (2000). “Principles for an emerging phonological system: a case study of early ASL acquisition,” in *Language Acquisition by Eye*, eds C. Chamberlain, J. P. Morford, and R. I. Mayberry (Mahwah, NJ: Lawrence Erlbaum Associates), 71–90.
- Maxwell, M. (1980). *Language Acquisition in a Deaf Child: The Interaction of Sign Variations, Speech, and Print Variations* (Unpublished doctoral dissertation). The University of Arizona, Tempe, AZ.
- Meier, R. P. (2006). “The form of early signs: explaining signing children's articulatory development,” in *Advances in the Sign Language Development of Deaf Children*, eds B. Schick, M. Marschark, and P. E. Spencer (New York, NY: Oxford University Press), 202–230.
- Meier, R. P., Cormier, K., and Quinto-Pozos, D. (2002). *Modality and Structure in Signed and Spoken Languages*. Cambridge: Cambridge University Press.
- Morgan, G., Barrett-Jones, S., and Stoneham, H. (2007). The first signs of language: phonological development in British Sign Language. *Appl. Psycholinguist.* 28, 3–22. doi: 10.1017/S0142716407070014
- Newport, E. L., and Meier, R. P. (1985). “The acquisition of American Sign Language,” in *The Crosslinguistic Study of Language Acquisition. Volume 1: The data*, ed D. I. Slobin (Hillsdale, NJ: Erlbaum), 881–938.
- Ohta, M. (1987). Cognitive disorders of infantile autism: a study employing the WISC, spatial relationship conceptualization, and gesture imitations. *J. Autism Dev. Disord.* 17, 45–62. doi: 10.1007/BF01487259
- Papadatou-Pastou, M., and Sáfár, A. (2016). Handedness prevalence in the deaf: meta-analyses. *Neurosci. Biobehav. Rev.* 60, 98–114. doi: 10.1016/j.neubiorev.2015.11.013
- Peterson, G. E., and Barney, H. L. (1952). Control methods for the study of vowels. *J. Acoust. Soc. Am.* 24, 175–184. doi: 10.1121/1.1906875
- Pierpaoli, C., Ferrante, L., Manzoni, T., and Fabri, M. (2014). Anatomical or mirror mode imitation? a behavioral approach. *Arch. Ital. Biol.* 152, 20–31.
- Pizer, G., Meier, R. P., and Shaw Points, K. (2011). “Child-directed signing as a linguistic register,” in *Formational Units in Sign Languages*, eds R. Channon and H. van der Hulst (Boston, MA: Walter de Gruyter), 65–86. doi: 10.1515/9781614510680.65
- Poizner, H., Klima, E. S., and Bellugi, U. (1990). *What the Hands Reveal About the Brain*. Cambridge, MA: MIT Press.
- Press, C., Ray, E., and Heyes, C. (2009). Imitation of lateralised body movements: doing it the hard way. *Laterality* 14, 515–527. doi: 10.1080/13576500802607990
- Rosen, R. (2004). Beginning L2 production errors in ASL lexical phonology: A cognitive phonology model. *Sign Lang. Linguist.* 7, 31–61. doi: 10.1075/sll.7.1.04beg
- Sandler, W., and Lillo-Martin, D. C. (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.
- Sharma, S. D. (2014). *Handedness and Cerebral Lateralisation: Looking at Signing, Fingerspelling and Gesture Skills in Deaf Signers and Signers with Stroke in British Sign Language (BSL)* (Unpublished doctoral dissertation). University College London, London, UK.
- Shield, A. (2010). *The Signing of Deaf Children With autism: Lexical Phonology and Perspective-Taking in the Visual-Spatial Modality* (Unpublished doctoral dissertation). University of Texas at Austin, Austin, TX.
- Shield, A. (2014). Preliminary findings of similarities and differences in the signed and spoken language of children with autism. *Semin. Speech Lang.* 35, 309–320. doi: 10.1055/s-0034-1389103
- Shield, A., Cooley, F., and Meier, R. P. (2017a). Sign language echolalia in deaf children with autism spectrum disorder. *J. Speech Lang. Hear. Res.* 60, 1–13. doi: 10.1044/2016_JSLHR-L-16-0292
- Shield, A., Knapke, K., Henry, M., Srinivasan, S., and Bhat, A. (2017b). Impaired praxis in gesture imitation by deaf children with autism spectrum disorder. *Autism Dev. Lang. Impairm.* 2, 1–14. doi: 10.1177/2396941517745674
- Shield, A., and Meier, R. P. (2012). Palm reversal errors in native-signing children with autism. *J. Commun. Disord.* 45, 439–454. doi: 10.1016/j.jcomdis.2012.08.004
- Shield, A., Meier, R. P., and Tager-Flusberg, H. (2015). The use of sign language pronouns by native-signing children with autism. *J. Autism Dev. Disord.* 45, 2128–2145. doi: 10.1007/s10803-015-2377-x
- Shield, A., Pyers, J., Martin, A., and Tager-Flusberg, H. (2016). Relations between language and cognition in native-signing children with autism spectrum disorder. *Autism Res.* 9, 1304–1315. doi: 10.1002/aur.1621

- Siedlecki, T., and Bonvillian, J. D. (1993). Location, handshape, and movement: Young children's acquisition of the formational aspects of American Sign Language. *Sign Lang. Stud.* 78, 31–52.
- Smith, I. M. (1998). Gesture imitation in autism I: Nonsymbolic postures and sequences. *Cogn. Neuropsychol.* 15, 747–770. doi: 10.1080/026432998381087
- Tager-Flusberg, H., and Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder: the neglected end of the spectrum. *Autism Res.* 6, 468–478. doi: 10.1002/aur.1329
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
- Watkins, F., and Thompson, R. L. (2017). The relationship between sign production and sign comprehension: what handedness reveals. *Cognition* 164, 144–149. doi: 10.1016/j.cognition.2017.03.019
- Whiten, A., and Brown, J. (1998). "Imitation and the reading of other minds: perspectives from the study of autism, normal children and non-human primates," in *Intersubjective Communication and Emotion in Early Ontogeny*, ed S. Bråten (Cambridge: Cambridge University Press), 260–280.
- Woll, B., Kyle, J. G., and Ackerman, J. (1988). Providing sign language models: strategies used by deaf mothers. *First Language* 8, 80.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Shield and Meier. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



When Speech Stops, Gesture Stops: Evidence From Developmental and Crosslinguistic Comparisons

Maria Graziano^{1*} and Marianne Gullberg^{1,2*}

¹ Lund University Humanities Lab, Lund University, Lund, Sweden, ² Centre for Languages and Literature, Lund University, Lund, Sweden

OPEN ACCESS

Edited by:

Guillaume Thierry,
Bangor University, United Kingdom

Reviewed by:

Pilar Prieto,
Institució Catalana de Recerca i
Estudis Avançats (ICREA), Spain
Katharina J. Rohlfing,
University of Paderborn, Germany

*Correspondence:

Maria Graziano
maria.graziano@humlab.lu.se
Marianne Gullberg
marianne.gullberg@ling.lu.se

Specialty section:

This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: 31 December 2017

Accepted: 15 May 2018

Published: 01 June 2018

Citation:

Graziano M and Gullberg M (2018)
When Speech Stops, Gesture Stops:
Evidence From Developmental
and Crosslinguistic Comparisons.
Front. Psychol. 9:879.
doi: 10.3389/fpsyg.2018.00879

There is plenty of evidence that speech and gesture form a tightly integrated system, as reflected in parallelisms in language production, comprehension, and development (McNeill, 1992; Kendon, 2004). Yet, it is a common assumption that speakers use gestures to compensate for their expressive difficulties, a notion found in developmental studies of both first and second language acquisition, and in theoretical proposals concerning the gesture-speech relationship. If gestures are compensatory, they should mainly occur in disfluent stretches of speech. However, the evidence is sparse and conflicting. This study extends previous studies and tests the putative compensatory role of gestures by comparing the gestural behavior in fluent vs. disfluent stretches of narratives by competent speakers in two languages (Dutch and Italian), and by language learners (children and adult L2 learners). The results reveal that (1) in all groups speakers overwhelmingly produce gestures during fluent speech and only rarely during disfluencies. However, L2 learners are significantly more likely to gesture in disfluency than the other groups; (2) in all groups gestures during disfluencies tend to be holds; (3) in all groups the rare gestures completed in disfluencies have both referential and pragmatic functions. Overall, the data strongly suggest that when speech stops, so does gesture. The findings constitute an important challenge to both gesture and language acquisition theories assuming a mainly (lexical) compensatory role for (referential) gestures. Instead, the results provide strong support for the notion that speech and gestures form an integrated system.

Keywords: gesture, speech production, language development, second language acquisition, crossmodal coordination

INTRODUCTION

In a seminal paper entitled *So you think gestures are non-verbal?* David McNeill challenged the then dominant view of gestures as a communicative frill of no consequence to our understanding of language and linguistic processing (McNeill, 1985). The paper listed arguments for why gestures are in fact verbal (i.e., linguistic), by highlighting their close relationship with spoken language in language development, in language break-down, and in language processing. He argued that speech and gesture develop in parallel in childhood, that the modalities break down together, and that they are processed in parallel in crossmodal information integration. There is now a substantial literature to support this view providing both behavioral and neurocognitive empirical

evidence to show that speech and gesture form an integrated mode of expression in production and comprehension (e.g., Kendon, 1980, 2004; McNeill, 1992, 2005; Willems and Hagoort, 2007 for overviews), in development (e.g., Capirci and Volterra, 2008; Colletta et al., 2015 for overviews), and across different spoken languages (Kita, 2009 for an overview). Yet, despite the evidence for such crossmodal integration, both empirical studies and theoretical proposals concerning the speech-gesture relationship often see gestures as having mainly a facilitating or compensatory function, helping speakers to overcome expressive difficulties (Gullberg, 1998, 2011 for overviews). However, the evidence concerning the precise link between speech break-down or disfluency and gestures remains contradictory. Therefore, the current study aims to examine the distribution of gestures relative to disfluencies in competent adult native speakers of two languages, and of language learners, both children and adults, in order to shed some light on the putative compensatory role of manual gestures, extending previous studies. In the following, we review the empirical and theoretical background to the study of disfluency in general, and to the temporal and functional relationship between speech and gesture specifically, including possible crosslinguistic differences, before turning to the current empirical study.

BACKGROUND

Despite ever-growing evidence for the integrated nature of speech and gesture, many empirical studies still view gestures as serving mainly a compensatory function. For example, in many studies of infants or very young children, gestures are described as behaviors preceding and preparing for language (Bates, 1979; Volterra et al., 1979; Liszkowski, 2008), paving the way for and predicting later linguistic development (e.g., Morford and Goldin-Meadow, 1992; Iverson et al., 1994; Capirci et al., 1996, 2005; Butcher and Goldin-Meadow, 2000; Özçalışkan and Goldin-Meadow, 2005; Pizzuto et al., 2005), and even facilitating access to the child lexicon (e.g., Pine et al., 2007). Gestures are thus generally implicitly described as having a facilitating function. In contrast, in adult second language acquisition and bilingualism studies, the compensatory view is explicit. Adult learners are often observed to be producing more gestures when speaking their second compared to their first language. This behavior is generally described as reflecting a compensatory effort to overcome lack of skill and fluency in the weaker language (Gullberg, 1998, 2011), or even as activating items in the mental lexicon (e.g., Nicoladis et al., 2007, 2009). Finally, studies of atypically developing or impaired populations also display a compensatory view of gesture. Children with Specific Language Impairment (SLI) or with Down syndrome show higher gesture rates than typically developing peers (e.g., Fex and Månsson, 1998; Stefanini et al., 2008), and so do aphasic patients, especially those with word retrieval impairments (e.g., Feyereisen, 1983; Hadar et al., 1998; Rose, 2006 for an overview). These higher gesture rates are all seen as evidence that gestures facilitate speaking or at least communicating.

Moreover, several theoretical accounts concerning the speech-gesture relationship also have compensatory foundations, revolving around how mainly referential¹ gestures, which convey information about referents' size, shape, movement or location, help speaking and thinking. For example, the Information Packaging Hypothesis (e.g., Alibali et al., 2000; Kita, 2000) suggests that referential gestures facilitate the conceptual planning of the spoken message, particularly of spatio-motoric concepts.² A recent expanded version, the Gesture-for-Conceptualization Hypothesis (Kita et al., 2017), proposes that speakers can activate, manipulate, package, and explore spatio-motoric information both for speaking and thinking through referential gestures. Although there is an underlying strand of compensatory thinking in these theories, their scope is very broad and the notion of compensation is not explicit. In contrast, the Lexical Retrieval Hypothesis (Krauss and Hadar, 1999; Krauss et al., 2000; Morsella and Krauss, 2005) is explicitly compensatory, suggesting that the main role of referential gestures is to facilitate lexical retrieval from the mental lexicon through crossmodal priming. In studies testing this theory, participants are often asked to name objects, or to provide words to a given definition, and in some cases are put in a tip-of-tongue state. These studies find that speakers produce more referential gestures when they speak about spatial content, and crucially, when they are searching for a word that is difficult to retrieve or that is unfamiliar (Butterworth and Hadar, 1989; Morrel-Samuels and Krauss, 1992; Rauscher et al., 1996; Krauss, 1998; Morsella and Krauss, 2005). More specifically, the claim is that word retrieval is more successful when participants gesture during the word search, that is, during the disfluency. Under the argument that gestures facilitate word retrieval, the temporal link between gesture production and disfluencies becomes crucial.

Disfluency and the Temporal Speech-Gesture Relationship

The vast literature on speech errors and disfluency in speech production has examined when and where in an utterance speakers interrupt speech (e.g., Maclay and Osgood, 1959; Goldman-Eisler, 1968; Hawkins, 1972; Beattie and Butterworth, 1979; Levelt, 1983, 1989; Clark, 1996 *inter multa alia*). They reveal that the beginning of a clause is a vulnerable site and that disfluencies also often occur before content words. In addition, these studies have also provided taxonomies of different types of disfluency markers (e.g., filled and unfilled pauses, interruptions, repetition, and lengthening). Studies have also shown that speakers prefer to self-correct (Schegloff et al., 1977), and favor fluency over accuracy in interaction, which means that they tend to interrupt speech not when the problem in encoding is detected, but rather when speakers are ready to produce a repair (Seyfeddinipur et al., 2008). Other studies indicate that filled pauses may have a signaling function much

¹Referential gestures are also known in the literature as representational, sometimes further labeled iconic/metaphoric gestures (e.g., McNeill, 1992). We will use the term referential gesture, following Kendon, 2004.

²A related suggestion is that gestures may relieve cognitive load although this is not specifically related to language (e.g., Goldin-Meadow et al., 2001; Hostetter and Sullivan, 2011; Cook et al., 2012).

like discourse markers (Clark and Fox Tree, 2002), and that both forms and distribution of such filled pauses are language-specific (e.g., Trofimovich and Baker, 2006; de Leeuw, 2007). In adult L2 learners, (dis-)fluency is discussed in terms of proficiency and (foreign) language skills (e.g., Poullisse, 1999; Schmid and Fägersten, 2010; De Jong et al., 2013; Bergmann et al., 2015).

Studies that specifically examine gesture production in relation to disfluency draw on some of these findings. Most studies investigate the temporal relationship between the gestural movement and disfluency markers. They present contradictory evidence both regarding the exact timing of the gesture relative to the disfluency, and the presence/absence of gesture. For example, Butterworth and Beattie (1978) found that gestures were as likely to begin during a silent pause as during speech. Ragsdale and Silvia (1982) instead reported that gestures could begin just before or simultaneously with non-fluent speech. However, in this study a wide range of movements was included (posture change, body shifts, foot, leg, head, and hand movements), making assessments specifically for manual gestures difficult. Generally, these early studies suggest that gestures tend to occur in the neighborhood of disfluencies. However, later studies have instead reported that speech and gesture stop at the same time. For instance, it has been shown that in stuttering populations the two modalities are interrupted together (Mayberry et al., 1998; Mayberry and Jaques, 2000). In other studies gestures are shown to stop even before speech stops (Seyfeddinipur and Kita, 2001; Seyfeddinipur, 2006), or to be totally absent during pauses and other disfluency phenomena (Christenfeld et al., 1991; Yasinnik et al., 2005). Further to this, there is some evidence that in adult L2 speakers' gestures are less frequent during disfluent than fluent speech (Gullberg, 1998). The evidence for how gestures and disfluency may be linked is thus mixed.

The explanations for the contradictory findings are likely to be methodological in nature. An obvious issue is that studies have focused on different kinds of movement involving various body parts (head, hands, feet, etc.), or manual gestures with particular functions such as referential gestures only versus looking at all gestural movements. This makes it difficult to assess comparability. Similarly, it is not always clear what kind of disfluency is involved (unfilled pauses only, or also filled pauses, repetitions, etc.). And most importantly, it is often unclear which part of the gestural movement is considered when the timing of a spoken disfluency and a gesture is compared: the whole gesture phrase (starting from the preparation and including the stroke and any post-stroke hold), or only the stroke/core movement phase, etc. (cf. Kendon, 1980, 2004). Claims about whether speech or gesture stops first, for example, must be very specific with regard to gesture phase or movement analyses (e.g., Seyfeddinipur and Kita, 2001; Seyfeddinipur, 2006). When more detail is provided, some studies find, for example, that it is specifically gesture holds (i.e., the momentary suspension of a movement en route) that tend to coincide with speech pauses (Yasinnik et al., 2005; Park-Doob, 2010), even in children aged nine (Esposito and Marinaro, 2007).

Disfluency and Gestural Function

In addition to timing, studies present mixed evidence concerning what gestural functions occur in disfluencies. As indicated, the theories and many studies have focused on referential gestures expressing referential content in disfluency. However, some of the earlier studies indicated the presence of different gestural functions by referring to 'break-down' gestures (Beattie and Butterworth, 1979 following Freedman, 1972). McNeill (1985, 1992) have subsequently labeled these 'butterworths' or 'conduit gestures', highlighting how gestures in break-downs often refer to the break-down itself, not to the content of speech. Gullberg (1998, 2011) has provided empirical support for this view, showing that if native and second language speakers gesture during disfluencies, they often produce gestures that comment on the breakdown itself but do not represent the referential content of the sought words. Many of these gestures involve continued wrist turning to expose palms (labeled meta-pragmatic, or 'thinking gestures' by Gullberg, 'cyclic gestures' by Ladewig, 2014), or palm up gestures directed toward the interlocutor. Kendon (2004) calls many of these gestures that do not express referential content for pragmatic gestures. On the whole, however, evidence for what functions gestures have in disfluency is scarce.

Disfluency and Crosslinguistic Comparisons

Relatedly, most studies concerned with gesture and disfluency are based on English production (except Italian in Esposito and Marinaro, 2007, and German in Seyfeddinipur and Kita, 2001; Seyfeddinipur, 2006). There are no direct crosslinguistic comparisons of the relationship between gesture and speech in disfluency. However, reports are found in the literature of differences in the distribution of gesture functions in speakers of different languages. For example, in a pioneering study Efron (1941/1972) observed that Italian immigrants in the United States produced more referential gestures than Yiddish-speaking immigrants, who instead tended to produce more pragmatic gestures. Similarly, Kendon (2004) observed a wider range of pragmatic gestures in Italian speakers than in British and American English speakers. Gullberg (1998) also observed that native Swedish speakers produced more referential gestures than native French speakers who instead produced more non-referential gestures (specifically beats). If gesture functions in disfluencies vary, then crosslinguistic preferences for referential or pragmatic gestures may interact with the kind of gestural behavior found in disfluency. However, gestures and disfluency has not been examined crosslinguistically, to our knowledge.

Intermediate Summary

In sum, previous studies provide inconsistent evidence on the precise temporal relationship between gestures and (dis-)fluency, presumably due to methodological differences. This in turn makes it difficult to assess theoretical proposals such as the compensatory Lexical Retrieval Hypothesis in contrast to the view of speech and gesture as an integrated system. Moreover, there is only scant evidence for how gestures are functionally

distributed during disfluent speech despite the latent relevance of gesture function to the theories about gesture and speech break-down. Further to this, direct crosslinguistic comparisons of speech disfluency and gesture are absent in the literature in spite of the potential importance of such comparisons for theoretical claims. Finally, data on language learners is scarce, looking specifically at disfluency rather than on general linguistic development in connection to gesture production. Therefore, to improve our understanding of whether speech and gestures form an integrated mode of expression or whether gestures mainly serve a compensatory or facilitating role in speech production, the current study aims to test the core predictions from the Lexical Retrieval Hypothesis, and examine the precise temporal and functional relationship between gestures and disfluencies in competent adult native speakers of two languages, and in language learners, children and adults.

CURRENT STUDY

The Lexical Retrieval Hypothesis predicts that (a) ongoing gestures should occur in stretches of *disfluent* compared to fluent speech if they are to help crossmodally prime lexical items; (b) that these gestures should have referential functions linking the gesture to the referential content of the lexical item sought. Further, assuming that language learners are more disfluent than competent speakers, we infer that the hypothesis would predict (c) that this state of affairs should hold especially for language learners. In contrast, the view of speech and gesture as an integrated system predicts that ongoing gestures should mainly occur in stretches of *fluent* speech compared to disfluent speech. It makes no predictions about gestural functions; however, previous observations suggest that ongoing strokes in disfluency may have a pragmatic rather than a referential function, commenting on the breakdown rather than reflecting the referential content of the sought lexical item. Finally, it predicts no differences between competent speakers and learners. Neither view makes predictions about crosslinguistic differences.

The current study addresses these issues and extends previous studies by comparing the gestural behavior during fluent and disfluent speech in (a) adult native speakers of Dutch vs. Italian; (b) child learners vs. adult competent speakers of Italian; and (c) adult Dutch second language learners of French vs. adult native Dutch speakers. We ask (1) whether speakers predominantly produce gestures with fluent or with disfluent speech; (2) whether gestures occurring with disfluencies tend to be ongoing strokes or holds; (3) whether ongoing strokes during disfluencies have referential or pragmatic functions; (4) and whether there are crosslinguistic differences between Dutch and Italian speakers.

Method

Participants

The analyses draw on four multimodal corpora consisting of narrative production (story retellings) in a dyadic, interactive setting. The corpora are based on the narratives of 66 participants divided over four groups (cf. **Table 1**): children learning Italian aged four, six, and nine ($n = 3 \times 11$, 22 female); adult Italian

TABLE 1 | Overview of participants.

	Mean age (year;month)	Age range (year; month)
Learners		
4-year-olds ($n = 11$; 6 f ¹)	4;7	4;1–5;4
6-year-olds ($n = 11$; 9 f)	6;8	6–7;8
9-year-olds ($n = 11$; 7 f)	9;2	8–10;9
Adult learners of L2 French ($n = 11$; 9 f) ²	20	19–22
Competent speakers		
Adult Italian L1 ($n = 11$; 7 f)	22	19–31
Adult Dutch L1 ($n = 11$; 9 f) ²	20	19–22

¹f, female.

²These are the same individuals.

native speakers ($n = 11$, 7 female); adult Dutch native speakers ($n = 11$, 9 female), who are also second language learners of French ($n = 11$, 9 female). The corpora thus consist of adult native speakers of two languages (Dutch, Italian) allowing for a crosslinguistic comparison of ‘competent’ speakers, and two types of learners (children, adults), allowing for a comparison of different types of learners (first vs. second language, L1 vs. L2).

Thirty-three Italian children were recruited in Naples ($n = 26$) and Rome ($n = 7$). The 11 Italian adults were university students recruited in Naples at the Università degli Studi di Napoli “L’Orientale”. The 11 Dutch adults were recruited at Radboud University, Nijmegen, Netherlands. They participated twice, speaking L1 Dutch on one occasion, and L2 French on the other. At the time of recording they had studied French as a foreign language for a minimum of 4 years, and had never lived in a French-speaking country. In some cases, 3 years had lapsed between their last contact with the language and the time of testing. They were all at a low to intermediate proficiency level. All participants signed a consent form; parents signed consent forms for the children.

Materials

All participants retold cartoon stories. Two different cartoons were used as stimuli. The Italian participants (children and adults) were shown a video entitled *Pingu’s family celebrates Christmas* (The Pygos Group, 1992), an episode lasting 90 s. The Dutch participants (native speakers and learners) were shown a printed wordless cartoon featuring three gnomes trying to solve a problem (cf. Gullberg, 2006). Since narrative content and structure is irrelevant to the analyses in this study, the use of different cartoons to elicit narrative production was deemed to be unproblematic.

Procedure

The Italian participants were presented with the cartoon on a laptop that was removed after viewing. Children were recorded in a familiar setting, either their home or at school. They retold the story to a familiar adult (a friend of the family or their teacher). The adult, who had also seen the cartoon, was instructed not to interrupt the child during the retelling, not to suggest parts of the story (even when the child missed them), but to provide feedback showing interest and participation to the interaction (i.e., *ah, uhu*,

I see, how nice). The Italian adults were recorded at university. Two participants were involved in each session: one person was asked to watch the cartoon and then to retell it to a friend who had not seen it. In order to make the Italian adult narratives comparable with those produced by the children, the listener was instructed to only listen to the story and to avoid interrupting the narrator, or to ask questions at the end of the story.

The Dutch participants were recorded at the Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, on two different occasions approximately a week apart: once in Dutch (the L1) and once in French (the L2). The order of the language/sessions was counterbalanced. The story was told to a confederate native speaker of the relevant language (Dutch for the L1 sessions, and French for the L2 sessions) who had not seen the cartoon. The interlocutor was instructed to ask clarification questions and provide feedback to create as naturalistic a session as possible.

Data Treatment and Coding

Data was transcribed and coded by frame-by-frame analysis of digital video in the annotation software ELAN (Wittenburg et al., 2006).

Speech

The retellings were transcribed using standard Dutch, French, and Italian orthography by native speakers. For the analyses presented here, all the L1 narratives (Dutch adults, Italian children and adults) were transcribed and analyzed in full (mean duration 2 min). Because the L2 narratives were considerably longer (mean duration 8 min), a selection was made of 2 min from the middle of the L2 recordings for transcription and analysis (see **Table 2**).

Speech was coded as fluent when no disfluency markers were present, or as disfluent when one of the following disfluency markers was present (boldface = disfluency marker):

- Filled pauses (*les deux **eh** nains* ‘the two uh dwarfs,’ D07L2);
- Unfilled pauses, minimum duration 200 ms, transcribed with (.) [*mettevano l’uovo (.) sopra* ‘they put the egg (.) on the top,’ ItCh12];
- Interruptions transcribed with apostrophe (*juste une **esc**’ escalier* ‘just one flight of st’ stairs,’ D21L2);
- Lengthenings, transcribed with colon (*alla fine **esce**: l’albero di Natale*; in the end it comes: the Christmas tree,’ ItCh24);

- Repetitions (*una palla **di di di** neve* ‘a ball of snow,’ ItCh16);
- Combinations of these categories with at least two different kinds of disfluencies appearing in immediate sequence [*il a une **eh** (.) **eh** image* ‘he has a eh (.) eh image,’ D01L2].

Importantly, only intra-clausal occurrences of disfluency were considered. That is, phenomena occurring at clause boundaries (as in example 1) or following discourse markers (2) were excluded.

(1) *i regali che hanno fatto ai genitori (.) nella terza scena troviamo che* (ItAd17) ‘the presents that they had made for the parents (.) in the third scene we find that’

(2) *allora (.) è vabbè l’inizio* (ItCh12) ‘well (.) uh well the beginning’

This selection was made to avoid over-estimating the amount of disfluencies. It is well-known that pauses often occur at clause- or utterance initial boundaries, and it is suggested that this is a consequence of the planning of the next clause (MacLay and Osgood, 1959; Hawkins, 1972, etc.). Moreover, it is also suggested that gestures are more likely to occur within than between clauses (cf. Beattie and Butterworth, 1979; McNeill, 1992, p. 94). In an examination of claims concerning speech and gestures in disfluency, instances of intra-clausal problems therefore seems like a better test bed where speech production has been launched and gestures are more likely to occur.

Twenty cases of repetition were excluded from analysis, since there were too few instances to perform further analysis. This procedure left 1,351 disfluencies for analysis. **Tables 3A,B** provide an overview of the aggregated and relative frequency distribution of fluent and disfluent stretches of speech across the groups, and the frequency of each of the disfluency markers, respectively.

Gestures

The gesture coding took the speech analysis as its departure point. First, for each fluent and disfluent stretch of speech, we coded for the presence or absence of a gesture. Second, gestures occurring with disfluent speech were further coded for their structural properties, that is, whether they were ongoing strokes or holds. Gestures were coded as ongoing when the stroke (i.e., the most effortful part of the gestural movement where the spatial excursion of the limb reaches its apex, cf. Kendon, 1980; McNeill, 1992; Seyfeddinipur, 2006) was being performed (**Figures 1B,C**). Gestures were coded as holds when there was a momentary suspension of movement, whether an interrupted or held preparation, or a post-stroke hold (**Figures 1D,E**; Kita et al., 1998). A total of 2,306 ongoing strokes, and 670 holds were identified. To give an overview of gestural activity in the data, we also computed mean gesture rate by word for each group, by dividing the total number of words (excluding interrupted words in disfluencies) with the total number of ongoing strokes per individual. We then computed the mean rate across each group. **Table 4** summarizes the distribution of ongoing strokes and mean gesture rate across groups to illustrate the properties of the sample.

Third, we coded all ongoing strokes (both in fluent and disfluent speech) for function. Following Kendon (2004), we

TABLE 2 | Overview of duration of retellings.

	Mean duration retellings (min:sec)	Analyzed
Learners		
4-year-olds (<i>n</i> = 11)	02:12	02:12
6-year-olds (<i>n</i> = 11)	02:33	02:33
9-year-olds (<i>n</i> = 11)	02:25	02:25
Adult learners of L2 French (<i>n</i> = 11)	08:23	02:00
Competent speakers		
Adult Italian L1 (<i>n</i> = 11)	01:51	01:51
Adult Dutch L1 (<i>n</i> = 11)	03:01	03:01

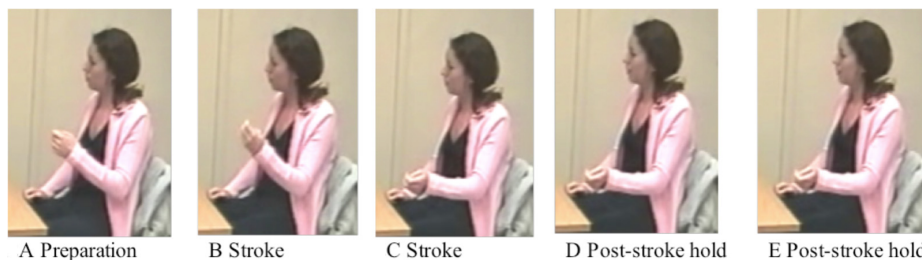
TABLE 3A | Number and mean proportion of fluent and disfluent stretches of speech across groups.

Learners	# Fluent stretches	# Disfluent stretches	M % disfluent stretches (SD)
4-year-olds	308	85	0.21 (0.08)
6-year-olds	495	155	0.23 (0.05)
9-year-olds	537	221	0.30 (0.06)
Adult learners of L2 French	471	395	0.46 (0.08)
Competent speakers			
Adult Italian L1	719	285	0.28 (0.08)
Adult Dutch L1	603	210	0.26 (0.05)

TABLE 3B | Number of types of disfluencies across groups.

Learners	Types of disfluencies				
	UP	FP	I	L	C
4-year-olds	39	4	20	11	11
6-year-olds	17	16	48	38	36
9-year-olds	32	23	57	75	34
Adult learners of L2 French	110	118	38	32	97
Competent speakers					
Adult Italian L1	10	32	18	204	21
Adult Dutch L1	48	100	4	16	42

UP, unfilled pause; FP, filled pause; I, interruption; L, lengthening; C, combination.

**FIGURE 1** | Example of gesture phases including ongoing stroke and post-stroke hold. (A) Preparation. (B) Stroke. (C) Stroke. (D) Post-stroke hold. (E) Post-stroke hold.

distinguished between referential and pragmatic functions. Gestures with a referential function (example in **Figure 2**) express semantic content through the depiction of referential

TABLE 4 | Frequency of gesture strokes and mean gesture rate/word across the groups.

Learners	# Gesture strokes	Mean gesture rate/word (SD)
4-year-olds	142	0.11 (0.06)
6-year-olds	325	0.13 (0.06)
9-year-olds	408	0.14 (0.05)
Adult learners of L2 French	392	0.29 (0.07)
Competent speakers		
Adult Italian L1	692	0.24 (0.07)
Adult Dutch L1	347	0.09 (0.03)
Total	2,306	

properties (e.g., size, shape, and action) or indexical properties (deictic gestures and pointing). Gestures with a pragmatic function (example in **Figure 3**), in contrast, convey part of “an utterance’s meaning that [is] not part of its referential meaning or propositional content” (Kendon, 2004, p. 158). In other words, pragmatic gestures do not express referential content but rather function like speech acts by commenting on the speaker’s spoken production. For this coding, we excluded those gestures that could not be determined as having either a referential or pragmatic function ($n = 35$ or 8% of the total number of gestures).

Finally, a new coder coded 10% of the data across all groups. We computed interrater reliability measures (Cohen’s kappa, cf. Hallgren, 2012) for the identification of disfluencies, and gestures, the coding of gestures as ongoing vs. holds, and gesture function as referential or pragmatic (**Table 5**).

Analyses

For all analyses, we make (a) a crosslinguistic comparison of competent adult native speakers of Dutch and Italian; (b) a



FIGURE 2 | Example of a referential gesture depicting fist fighting.

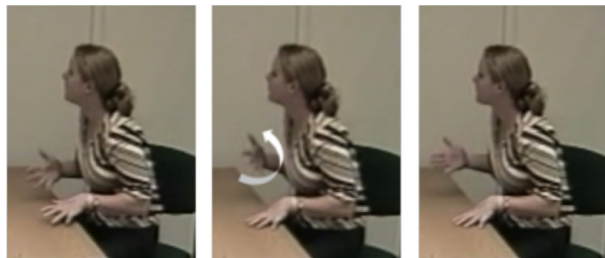


FIGURE 3 | Example of a pragmatic gesture.

developmental comparison of three Italian child groups and adult Italian speakers; (c) a developmental comparison between competent adult native speakers of Dutch and adult Dutch L2 learners of French.

For the statistical analyses we used the *glmerMod* package in R, version 0.98.953 (R Core Team, 2014) to perform Generalized Linear Mixed-effects Models (GLMMs) with random intercepts for participants and items (Baayen, 2008; Baayen et al., 2008). Models were fit using maximum likelihood (Laplace approximation) [*glmerMod*], binomial family (logit), since the dependent variable outcome throughout was binary. All analyses were run on raw numbers, but for ease of exposition figures show mean proportions.

TABLE 5 | Interrater reliability measures, Cohen's kappa.

Group	Data	Observations	Raters	Kappa
Italian_L1	Speech disfluency	263	2	0.936
Italian_L1	Gesture	357	2	0.955
Italian_L1	Ongoing stroke/hold	267	2	0.867
Italian_L1	Gesture function	223	2	0.992
Dutch_L1	Speech disfluency	82	2	0.868
Dutch_L1	Gesture	90	2	0.937
Dutch_L1	Ongoing stroke/hold	37	2	0.874
Dutch_L1	Gesture function	35	2	0.968
Adult_L2	Speech disfluency	59	2	0.984
Adult_L2	Gesture	118	2	0.920
Adult_L2	Ongoing stroke/hold	116	2	0.987
Adult_L2	Gesture function	72	2	0.975

RESULTS

Gestures With Disfluent vs. Fluent Speech

Figure 4 presents the mean proportion of ongoing strokes occurring with disfluent and fluent speech, respectively, comparing adult native Dutch and Italian speakers (**Figure 4A**), Italian 4-, 6-, and 9-year-olds and adult Italian speakers (**Figure 4B**), and adult native Dutch speakers and adult Dutch learners of L2 French (**Figure 4C**). **Table 6** presents the output from three GLMMs on the likelihood of gestures occurring with disfluent speech across groups, again, first examining adult native Dutch and Italian speakers; then Italian 4-, 6-, and 9-year-olds and adult Italian speakers; and finally, adult native Dutch speakers and adult Dutch learners of L2 French. Participants and items were always random factors, and group (Dutch/Italian; 4-/6-/9-year-old/adult Italian; L1/L2) and speech (disfluent/fluent) fixed main effects.

The results indicate that in all groups there was a main effect of speech type such that gestures were significantly more likely to occur with fluent than disfluent speech (adult Dutch/adult Italian, *Est.* = 2.491, *z* = 17.114, *p* < 0.001; Italian 4-/6-/9-year-olds/adults, *Est.* = 2.2942, *z* = 20.253, *p* < 0.001; and L1 Dutch/L2 French, *Est.* = 2.1997, *z* = 9.512, *p* < 0.001). In addition, the results reveal a shift over the course of child development, with Italian adults (*Est.* = 1.8585, *z* = 5.291, *p* < 0.001) and 9-year-olds (*Est.* = 0.885, *z* = 2.539, *p* < 0.05) differing from 4-year-olds who do not differ from 6-year-olds. Furthermore, for L2 speakers there is an interaction with speech type such that L2 speakers are significantly more likely than L1 speakers to produce gestures with disfluent speech (*Est.* = -0.8697, *z* = -3.176, *p* < 0.01).

The following examples illustrate the main pattern of absence of gestures during disfluencies. We follow Kendon (2004) in transcribing gestures: | = gesture phrase/unit boundaries; ~ = preparation phase; ** = stroke; underline = hold; - = recovery.

(3) adult Dutch native speaker D25L1

en t' derde mannetje die gaat er dus vandoor met ehm (.) de ladder

|*****|*****|
'and the third little man he just goes ahead with uh'

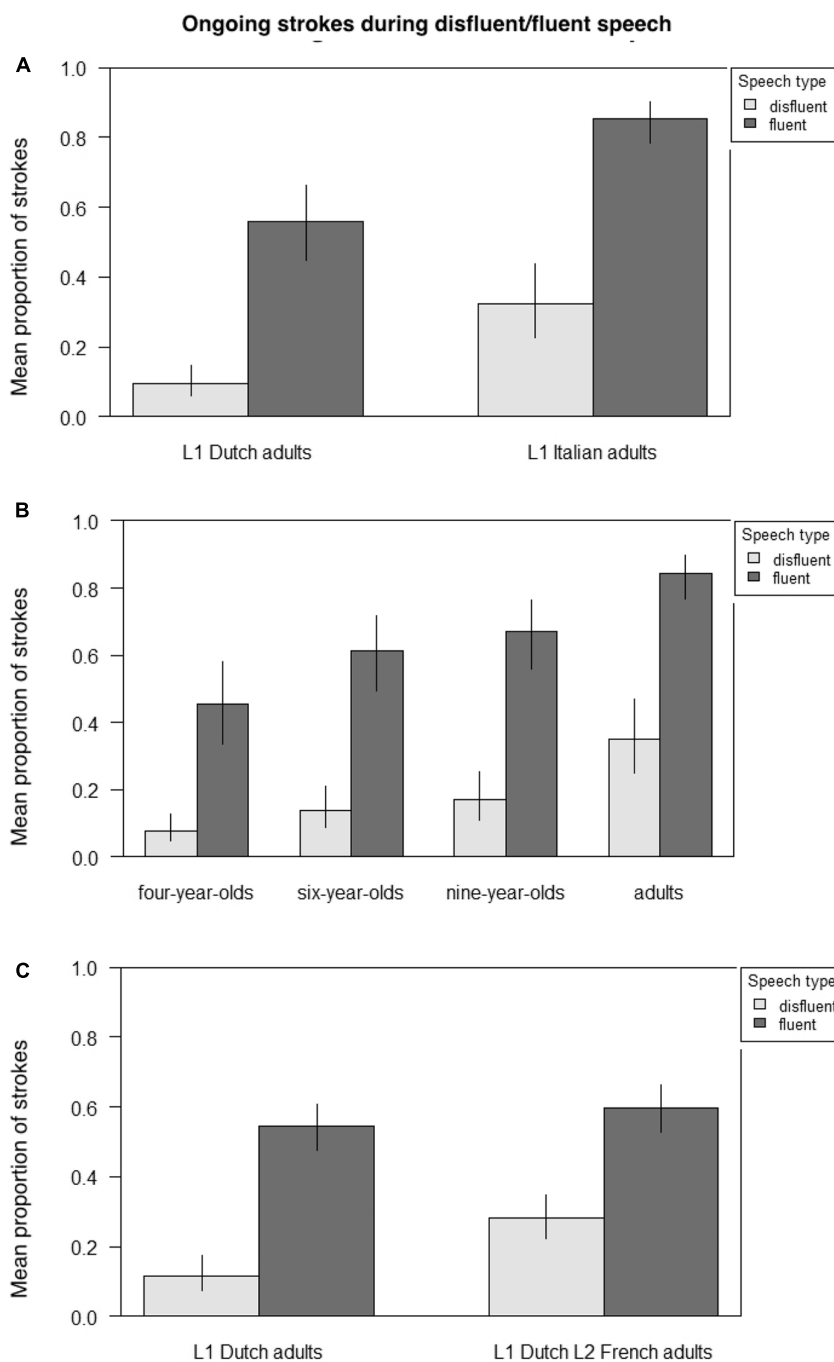


FIGURE 4 | Mean proportion of ongoing strokes during disfluent/ fluent speech across groups. **(A)** Adult native Dutch vs. Italian speakers. **(B)** Italian children aged 4, 6, and 9 vs. Italian adult speakers. **(C)** Adult native Dutch speakers vs. adult Dutch learners of L2 French.

In (3) a Dutch native speaker says *en t' derdre mannetje die gaat er dus vandoor met* 'and the third little man he just goes ahead with' producing two gestures. The first is a referential gesture where both hands have a tight grip handshape moving rightward, as if holding something and moving it. The second gesture is a pragmatic gesture where the both hands are twisted at the wrist to reveal palms up. When she then becomes disfluent, starting with

a filled pause followed by a long silence, she drops both hands to the lap.

(4) adult Italian native speaker (ItAd05)
il padre fuori l'igloo che: che: appunto addobba

|*****|*****|

'the father outside the igloo that: that: in fact decorate'

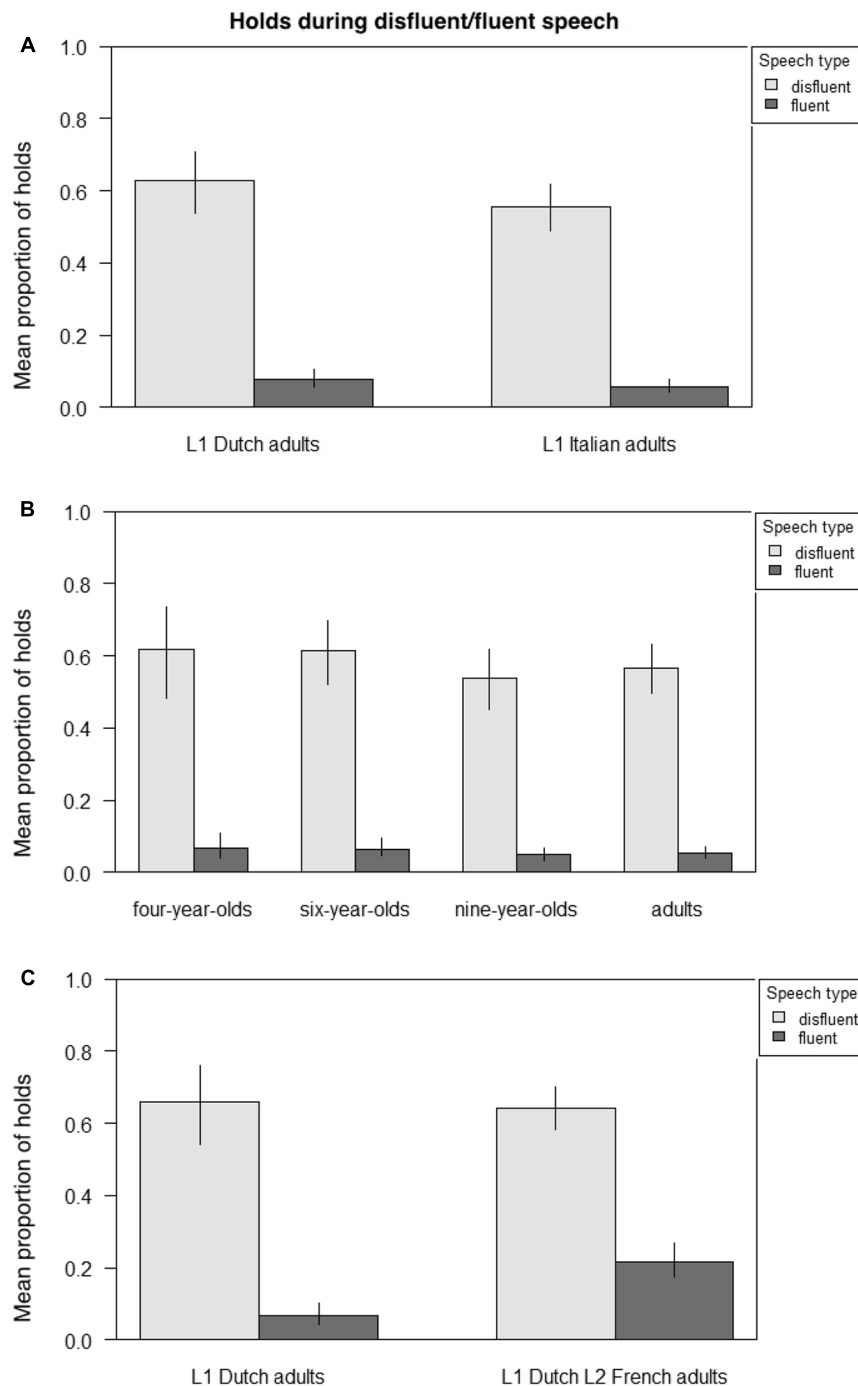


FIGURE 5 | Mean proportion of gestural holds during disfluent/fluently speech across groups. **(A)** Adult native Dutch vs. Italian speakers. **(B)** Italian children aged 4, 6, and 9 vs. Italian adult speakers. **(C)** Adult native Dutch speakers vs. adult Dutch learners of L2 French.

Gesture Functions in Disfluent Speech

Figure 6 presents the mean proportion of gestures with a pragmatic function across fluent and disfluent stretches of speech, respectively, comparing adult native Dutch and Italian speakers (Figure 6A), Italian 4-, 6-, and 9-year-olds and adult Italian speakers (6B), and adult native Dutch speakers (6B), and adult native Dutch speakers and adult Dutch

learners of L2 French (6C). Table 8 presents the output from three GLMMs on the likelihood of pragmatic gestures occurring with disfluent speech across groups, again, first examining adult native Dutch and Italian speakers; then Italian 4-, 6-, and 9-year-olds and adult Italian speakers; and finally, adult native Dutch speakers in L1 and in L2 French. Participants and items were

TABLE 7 | Summary of Generalized Linear Mixed Models testing whether gestural holds occur mostly with disfluent vs. fluent speech across groups.

	Est.	SE	z
Adult Dutch/Adult Italian (gesture2 ~ group+speech2+1(1 participant))			
Intercept	-0.5206	0.1877	-2.774**
groupItalianAdu	0.2993	0.1919	1.559
speechFluent	3.0070	0.1815	16.570***
Italian 4-/6-/9-year-olds/adults (gesture2 ~ group+speech2+1(1 participant))			
Intercept	-0.4770	0.2820	-1.692
groupItalian6ys	0.0140	0.3194	0.044
groupItalian9ys	0.3314	0.3127	1.060
groupItalianAdu	0.2154	0.2959	0.728
speechFluent	3.1174	0.1542	20.211***
Adult Dutch L1 vs. Adult L2 French (gesture2 ~ group*speech+1(1 participant))†			
Intercept	-0.6598	0.2542	-2.569**
groupL2	0.0735	0.2742	0.268
speechFluent	3.2821	0.3262	10.062***
groupL2:speechFluent	-1.4160	0.3699	-3.828***

p-values: ***0.001, **0.01, * 0.05.

†The model with the interaction term better explained the data and was therefore selected, $\chi^2(1) = 15.519$, $p < 0.001$.

always random factors, and group (Dutch/Italian; 4-/6-/9-year-old/adult Italian; L1/L2) and speech (disfluent/fluent) fixed main effects.

The results indicate that in no group were pragmatic gestures more likely to occur with disfluent than fluent speech despite numerical trends in some groups. However, there was a crosslinguistic difference in that Italian speakers were more likely to produce pragmatic gestures with fluent speech than adult Dutch speakers (Est. = -2.1988, $z = -5.261$, $p < 0.001$). There was also a developmental effect in that Italian 9-year-olds (Est. = -1.3441, $z = -2.714$, $p < 0.01$) and adults (Est. = -4.266, $z = -4.600$, $p < 0.001$) were more likely to produce pragmatic gestures with fluent speech than 4- and 6-year-olds, who did not differ. Finally, adult L2 speakers were significantly more likely to produce pragmatic gestures with fluent L2 speech than L1 speech (Est. = -1.4160, $z = -3.828$, $p < 0.001$).

Examples (8) and (9) illustrate the occurrence of pragmatic gestures during disfluencies.

(9) Italian child learner (ItCh31)

con matterello stava: (.) *stendendo la sfoglia per fare dei biscotti*

|**-.| |*****|

'with the rolling pin was: (.) stretching out the pastry to make cookies'

In (9), an Italian 9-year-old hesitates and produces a gesture with a pragmatic function during the unfilled pause (.) with the right open hand rotated up and down twice. Once speech resumes, he continues to produce a referential gesture that represents the stretching out of the pastry with both hands.

(10) adult L2 learner of French (D21L2) <> = whispering

ilest eh (.) <putting> eh (.) *le maison est*

|**||**| |**|| |~*****-.|

'he is eh (.) <putting> eh (.) the house is'

In (10), an L2 speaker produces a string of gestures with pragmatic functions during a long disfluent stretch, tapping her fingers with both hands on the table. These gestures are accompanied by averted gaze and a thinking face (cf. Goodwin and Goodwin, 1986; Gullberg, 2011). When she resumes speech saying *le maison* 'the house,' she simultaneously produces a gesture with a referential function, fingers tracing a square.

A final example (10) illustrates how an onstroking stroke with a referential function is produced during a disfluency by a L2 speaker (L2 = L2 speaker; NS = native speaker interlocutor).

(11) adult L2 learner of French (D07L2)

L2: *ils sont (.) très eh (.)*
|**

NS: *en colère*

-.-.|

L2: *en colère et (.)*

|~****

NS: *ils se battent*

*****-.|

L2: *oui oui*

L2: 'they are (.) very uhm (.)

NS: angry

L2: angry and (.)

NS: they fight

L2: yes yes'

In the sequence in (11), after the L2 speaker initiates a fluent stretch, *ils sont très* 'they are very,' she becomes disfluent. In the second unfilled pause, she produces a gesture with a referential function representing the act of fighting with both fists moving around each other in a circle (cf. Figure 2). She shifts her gaze to the native interlocutor who offers a first solution, *en colère* 'angry' while the learner drops her hands. The L2 speaker repeats this phrase but is not satisfied, so she repeats the gesture in a third unfilled pause, again with gaze shifted to the native speaker. The learner's gesture has gone into a hold and is held while the native speaker suggests *ils se battent* 'they fight.' The learner accepts this suggestion, drops her hands, and confirms, *oui oui* 'yes yes,' nodding. The referential 'fighting gesture' is thus used to elicit the lexical item from the interlocutor (cf. Gullberg, 1998, 2011).

DISCUSSION

This study examined the putative compensatory role of gestures by investigating their distribution, temporal, and functional properties relative to speech disfluencies in speakers of two different languages (Dutch and Italian), and with different degrees of linguistic expertise (child and adult language learners). The key findings can be summarized in four points. First, in all groups, speakers' gesture production differs in fluent and disfluent stretches of speech, such that gestures overwhelmingly occur with fluent speech. Adult L2 speakers are more likely than anyone else to gesture also during disfluent speech. Second, in all groups gestures tend to be held during disfluent speech, not to be ongoing strokes. Third, the small number of ongoing gestures during

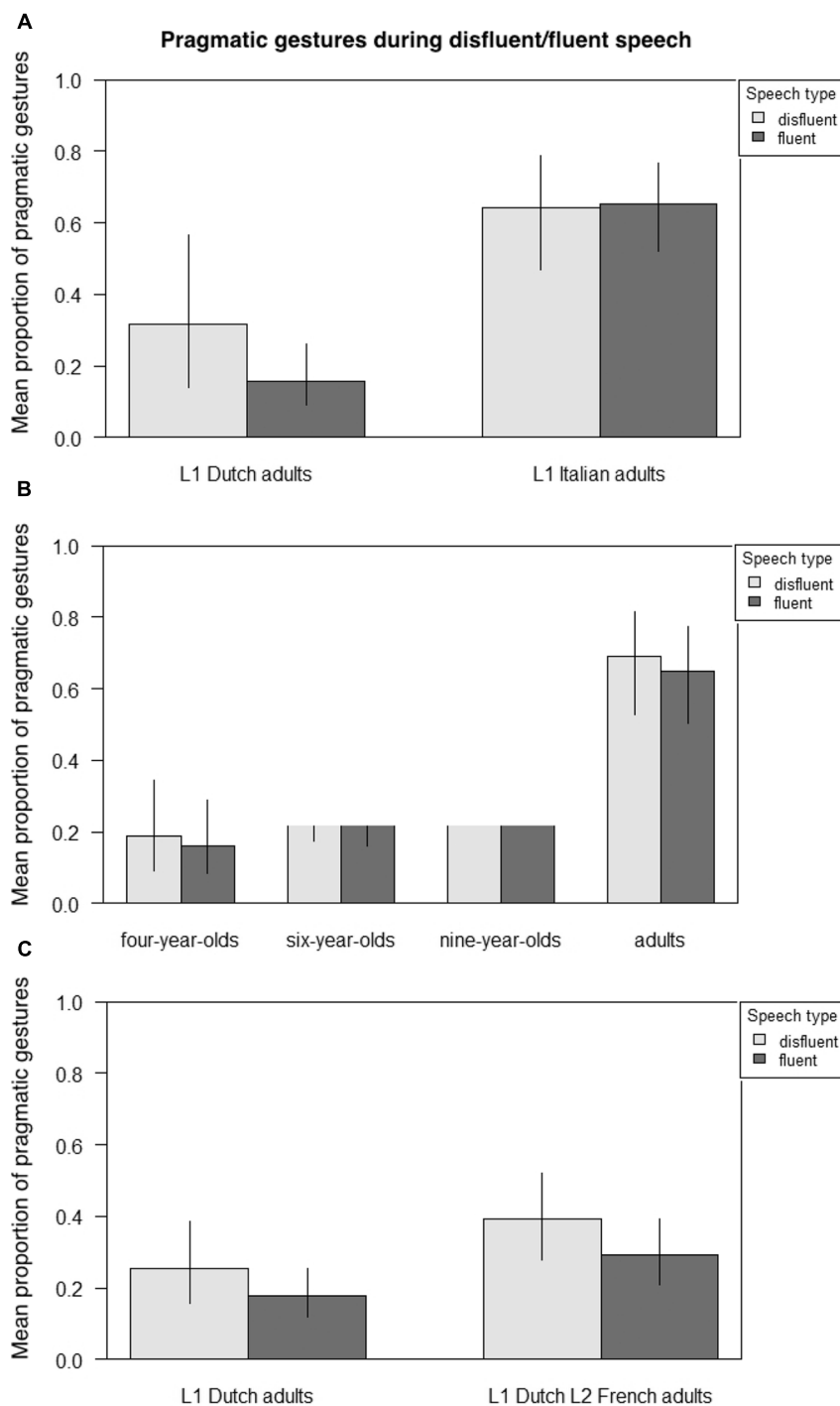


FIGURE 6 | Mean proportion of pragmatic gestures during disfluent/fluent speech across groups. **(A)** Adult native Dutch vs. Italian speakers. **(B)** Italian children aged four, six and nine vs. Italian adult speakers. **(C)** Adult native Dutch speakers vs. adult Dutch learners of L2 French.

disfluency display both pragmatic and referential functions. Adult L2 learners are more likely than anyone else to produce referential gestures during disfluency. Fourth, there are no crosslinguistic differences in gestural behavior during disfluencies. We only find a crosslinguistic difference in the

production of pragmatic gestures during *fluent* stretches, with Italian adults producing more such gestures than Dutch adults and Italian children.

The overwhelming tendency for gestures to occur with fluent rather than disfluent speech does not support the first prediction

TABLE 8 | Summary of Generalized Linear Mixed Models testing whether pragmatic gestures occur mostly with disfluent vs. fluent speech across groups.

	Est.	SE	z
Adult Dutch/Adult Italian (gestfunction ~ group+speech+1 (1 participant))			
Intercept	1.4438	0.3754	3.846***
groupItalianAdu	-2.1988	0.4179	-5.261***
speechFluent	0.1424	0.2363	0.603
Italian 4-/6-/9-year-olds/adults (gestfunction ~ group+speech+1 (1 participant))			
Intercept	1.4698	0.4204	3.496***
groupItalian6ys	-0.6383	0.5025	-1.270
groupItalian9ys	-1.3441	0.4953	-2.714**
groupItalianAdu	-2.2660	0.4927	-4.599***
speechFluent	0.1763	0.1859	0.949
Adult Dutch L1 vs. Adult L2 French (gestfunction ~ group+speech+1 (1 participant))			
Intercept	1.0786	0.3109	3.469***
groupL2	-0.6455	0.1948	-3.314***
speechFluent	0.4498	0.2309	1.948

p-values:***0.001, **0.01, * 0.05.

by the Lexical Retrieval Hypothesis to the effect that, if gestures facilitate lexical retrieval, they should occur more frequently during speech disfluencies. Instead, the results suggest a very tight link between fluent speech and gesture production, supporting the notion that speech and gesture form an integrated or co-orchestrated system in speech production (e.g., McNeill, 1992; Clark, 1996; Kendon, 2004). The strikingly similar patterns found across speakers of different languages and across competent and learning language users alike support this notion quite forcefully.

The finding that any gestural activity found during speech disfluencies is mostly held or suspended in all groups similarly further reinforces the view of an integrated speech-gesture system. All speakers, children and adults, competent or learners, either interrupt an ongoing gesture when speech is interrupted (i.e., they stop or hold the preparation) or they freeze it (i.e., produce a post-stroke hold). That is, when speech stops, so does gesture. This finding is in line with and extends previous studies (e.g., Mayberry and Jaques, 2000; Seyfeddinipur and Kita, 2001; Yasinnik et al., 2005; Esposito and Marinaro, 2007), and provides supplementary evidence that holds or gesture suspensions tend to coincide with disfluency markers. It is also in line with McNeill's suggestion of parallel break-downs (McNeill, 1985). These speaker-directed perspectives are complemented by findings on the functions of holds in interaction, which are relevant since the narratives analyzed here are interactive. For example, in seminal work Duncan (1972) showed that holds and 'relaxation' of tensed hands consistently occurred at the ends of turns in conversation thus signaling the end of a turn. When they linger after the turn, they have often been treated as cues to elicit a response from the interlocutor (Bavelas, 1994; Sikveland and Ogden, 2012; Cibulka, 2016, inter al.). Park-Doob (2010, p. 1) demonstrates that holds can "support continued expressiveness and interpretability," that is they can indicate that the concept

presented through the gesture is still active, thus allowing an interlocutor to draw information from a suspended gesture. Similarly, Cibulka (2016) reports that holds can be deliberately inserted in repair sequences to indicate that an entire utterance is momentarily suspended. Such functional analyses of holds in interaction are not in contradiction to the current findings concerning the speech production process. Instead, they provide a window on the multi-functionality of gestures in general and suspensions/holds in particular, whereby both speech and gesture production processes are subject to multiple influences in interaction (cf. Kendon, 2004).

Turning to gestural functions during disfluency, all groups produced not only referential but also pragmatic gestures in the small number of ongoing strokes found during disfluencies. Again, this result does not support the second prediction by the Lexical Retrieval Hypothesis, according to which we should expect referential gestures during disfluencies activating lexical items. As in the examples provided, the pragmatic gestures performed during disfluencies are not related to lexical content but rather to aspects of difficult interaction arising from the disfluencies both in adults and children (cf. Graziano, 2014a,b for similar findings on children). These gestures, often performed with a repeated oscillation of the open hand through wrist rotation or by tapping the fingers on a surface, provide a metalinguistic comment on the communication breakdowns, signaling that there is a problem in the speech production or that the speaker is engaging in a word search. Stam and Tellier (2017) classify word searching gestures as production oriented. This certainly tallies with these findings. However, although these gestures clearly indicate a production difficulty, they equally clearly have the potential to serve an interactive function (cf. Bavelas et al., 1992), indicating, for example, that the speaker is holding the floor. The averted gaze and the 'thinking face' (Goodwin and Goodwin, 1986) that often accompanies these gestures, suggest a strong floor-holding component.

Learners, both children and adults, overall revealed the same patterns as competent speakers, and there were no crosslinguistic differences in disfluencies. These findings highlight that the integrated behavior is pervasive. That said, the adult L2 speakers differed most from other groups both in speech and gesture. Although they overall pattern in the same way as the other groups, L2 speakers are more likely than native speakers to produce (ongoing and referential) gestures with disfluent speech. Although this result seems to support the predictions by the Lexical Retrieval Hypothesis, it is important to qualify the finding. First, it is not the dominant pattern even for L2 speakers. Second, ongoing strokes in disfluency have both pragmatic and referential functions. The pragmatic functions do not relate to lexical content, so cannot support lexical retrieval. Third, and most importantly, when referential gestures are produced during disfluencies, they tend to occur in specific contexts, illustrated by example (11). Here the L2 speaker seems to produce referential gestures strategically to elicit lexical help from the interlocutor – not from herself. In performing the 'fighting' gesture (cf. Figure 2) in silence, the L2 speaker certainly represents the concept she has trouble expressing,

but she also uses the referential dimension of the gesture in combination with the direct gaze to the interlocutor with a pragmatic aim, namely to request help from the interlocutor, who does indeed provide a linguistic label for the gesture. Such sequences are relatively common in face-to-face interaction between L2 and native speakers (cf. Gullberg, 1998, 2011). There is further support for the crucial interactive aspect of such behavior. Holler et al. (2013) have shown that the communicative situation affects the rate of referential gestures in disfluency. During non-fluent speech, native speakers tend to produce more referential gestures during tip-of-the-tongue states when facing interlocutors than when they cannot see them or when they speak to a recorder. Overall, such patterns of production of referential gestures in disfluencies support Kendon's (2004) claim that gestures, depending on the context, can have multiple functions at the same time; namely, in this case, referential and pragmatic/interactive. Obviously, this is not to say that referential gestures are never produced instead of lexical items or never ease their production. But we do claim that this cannot be considered the main function of gestures, not even for L2 speakers.

A further result from the L2 speakers is that they rather surprisingly produce more holds with *fluent* speech than anyone else. One possible reason for this is that the L2 speakers under study really are beginners with low levels of proficiency. They are therefore highly disfluent. In fact, they are so disfluent that their 'fluent' stretches of speech tend to be very short, consisting only of one or two words, and to be 'inserted' between disfluencies, rather than the other way around. Examples (6) and (9) illustrate this quite clearly. In such situations, suspensions or holds from a disfluency can 'spill over' to the fluent part of an utterance. On the whole, then, L2 speakers display more of everything than the other groups – they are more disfluent than any other group, but their predominant pattern of no gesture or hold in disfluency is the same as for all. They also produce more ongoing strokes with referential functions in disfluencies than anyone else. This is presumably a reflection of the fact that they may have a communicative intention ready in their first language which they cannot express lexically in the second language. Their referential gesture can thus reflect a lexical notion in the L1 when they decide to use the gesture to elicit help from an interlocutor. But if the word is not known in the L2, then no amount of gesturing can activate it.

It is important to acknowledge that the Lexical Retrieval Hypothesis makes predictions specifically concerning lexical difficulties in the domain of spatial language, assuming that referential gestures will crossmodally prime spatial vocabulary. The current analyses have not taken the specifics of lexical information into account, but rather applied a global analysis to all intra-clausal disfluencies. Partly, this is because we have conducted a corpus analysis on naturalistically occurring disfluencies in narrative corpora. In such contexts, it is not always easy to know whether the sought word is spatial or not, nor whether the resolution is even related to the original lexical problem (cf. Seyfeddinipur, 2006 for similar comments). However, it seems unlikely that the overwhelmingly clear patterns found in the four corpora analyzed would change for spatial

language specifically. That said, an experimental study could be undertaken inducing disfluency and targeting specific semantic domains to see whether the type of analysis performed here would yield similar results. This would also address other drawbacks with the corpus analysis such as differing elicitation methods across corpora both as regards stimulus materials (printed/video) and common ground (whether interlocutors also saw the stimuli or not). Both differences may have affected overall gesture rate, for example, and although gesture rate was not of interest *per se* in this study, it may have influenced the sample size.

The current results provide no or little support for the Lexical Retrieval Hypothesis proposing that ongoing referential gestures in disfluencies help speech production. But what about the ongoing pragmatic, or rather non-referential, gestures? Following other authors, we have suggested that these gestures comment on the break-downs in interactive settings. However, suggestions are found in the literature to the effect that non-referential gestures may serve a speaker-directed purpose, helping to stimulate and focus attention thus keeping "communicative speech "on course"" (e.g., Grand et al., 1977, p. 499; cf. Stam and Tellier, 2017). Admittedly, many findings are linked to the study of populations with psychiatric conditions, but they open potential new avenues of exploration.

CONCLUSION

Overall, the results from the present study suggest a very tight link between fluent speech and gesture production, providing strong support for the notion that speech and gestures form a tightly integrated or co-orchestrated system, with similar properties across languages and speakers' skills. The findings constitute an important challenge for gesture theories assuming a mainly (lexical) compensatory role for (referential) gestures. Moreover, the observation that gestures that do accompany disfluencies have both pragmatic and referential functions raises further important challenges for gesture theories which have hitherto been based on subsets of gestures (referential) and solely on adult, competent, fluent speakers. The findings are also challenging for theories of language acquisition that tend to view gestures mainly as a (lexical) crutch. Perhaps most importantly, the findings are a challenge for mono-modal theories of language who look only to (written forms of) spoken or signed language, ignoring gestures as irrelevant. The data strongly suggest that when speech stops, so does gesture across languages, across age, and across types of learners. Speech disfluency is generally mirrored by gesture disfluency. To us, this suggests that gesture production is part and parcel of language production, and therefore worthy of linguistic theorizing more broadly.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Regional Ethical Review Board at Lund University with written informed consent from all

subjects (note that the data were collected while the authors were employed in the Netherlands and Italy, but that the Swedish board has reviewed the protocol). All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

We gratefully acknowledge financial support from the Erik Philip-Sørensen Foundation, the European Science Foundation (Short Term Scientific Mission – COST Action

2102, n. 17), the Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands, and the Swedish Research Council (Vetenskapsrådet; A0667401).

ACKNOWLEDGMENTS

We thank two reviewers for very helpful comments on a previous version of the manuscript; research assistants Josine Greidanus for help with statistical analysis and Dutch transcription, Nicolas Femia, Frida Splendido, and Wanda Jakobsen for reliability coding. We also express our thanks to Prof. Amneris Roselli and Mrs. Pina Ciompi (Università degli Studi di Napoli “L’Orientale”) for hosting and supporting us during data collection in Naples. We also gratefully acknowledge support from Lund University Humanities Lab, Lund University, Sweden.

REFERENCES

- Alibali, M. W., Kita, S., and Young, A. J. (2000). Gesture and the process of speech production: we think, therefore, we gesture. *Lang. Cogn. Process.* 15, 593–613. doi: 10.1080/016909600750040571
- Baayen, R. H. (2008). *Analyzing Linguistic Data. A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511801686
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005
- Bates, E. (1979). *The Emergence of Symbols: Cognition and Communication in Infancy*. New York, NY: Academic Press.
- Bavelas, J. B. (1994). Gestures as part of speech: methodological implications. *Res. Lang. Soc. Interact.* 27, 201–221. doi: 10.1207/s15327973rlsi2703_3
- Bavelas, J. B., Chovil, N., Lawrie, D. A., and Wade, A. (1992). Interactive gestures. *Dis. Process.* 19, 163–194. doi: 10.1080/01638539209544823
- Beattie, G., and Butterworth, B. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Lang. Speech* 22, 201–211. doi: 10.1177/002383097902200301
- Bergmann, C., Sprenger, S. A., and Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychol.* 161, 25–35. doi: 10.1016/j.actpsy.2015.07.015
- Butcher, C., and Goldin-Meadow, S. (2000). “Gesture and the transition from one- to two-word speech: when hand and mouth come together,” in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 235–257.
- Butterworth, B., and Beattie, G. (1978). “Gesture and silence as indicators of planning in speech,” in *Recent Advances in the Psychology of Language: Formal and Experimental Approaches*, eds R. Campbell and P. T. Smith (London: Plenum), 347–360.
- Butterworth, B., and Hadar, U. (1989). Gesture, speech, and computational stages: a reply to McNeill. *Psychol. Rev.* 96, 168–174. doi: 10.1037/0033-295X.96.1.168
- Capirci, O., Contaldo, A., Caselli, M. C., and Volterra, V. (2005). From action to language through gesture: a longitudinal perspective. *Gesture* 5, 155–177. doi: 10.1075/gest.5.1-2.12cap
- Capirci, O., Iverson, J. M., Pizzuto, E., and Volterra, V. (1996). Gestures and words during the transition to two-word speech. *J. Child Lang.* 23, 645–673. doi: 10.1017/S0305000900008989
- Capirci, O., and Volterra, V. (2008). Gesture and speech. The emergence and development of a strong and changing partnership. *Gesture* 8, 22–44. doi: 10.1075/gest.8.1.04cap
- Christenfeld, N., Schachter, S., and Bilous, F. (1991). Filled pauses and gestures: it's not coincidence. *J. Psycholinguist. Res.* 20, 1–10. doi: 10.1007/BF01076916
- Cibulka, P. (2016). How to do things with holds. *Sign Lang. Stud.* 16, 447–472. doi: 10.1353/sls.2016.0015
- Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620539
- Clark, H. H., and Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition* 84, 73–111. doi: 10.1016/S0010-0277(02)00017-3
- Colletta, J.-M., Guidetti, M., Capirci, O., Cristilli, C., Demire, O. E., Kunene, R. N., et al. (2015). Effects of age and language on co-speech gesture production: an investigation of French, American, and Italian children's narratives. *J. Child Lang.* 42, 122–145. doi: 10.1017/S0305000913000585
- Cook, S. W., Yip, T. K., and Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Lang. Cogn. Process.* 27, 594–610. doi: 10.1080/01690965.2011.567074
- De Jong, N., Steinel, M., Florijn, A., Schoonen, R., and Hulstijn, J. (2013). Linguistic skills and speaking fluency in a second language. *Appl. Psycholinguist.* 34, 893–916. doi: 10.1017/S0142716412000069
- de Leeuw, E. (2007). Hesitation markers in English, German, and Dutch. *J. Germanic Linguist.* 19, 85–114. doi: 10.1017/S1470542707000049
- Duncan, S. J. (1972). Some signals and rules for taking speaking turns in conversation. *J. Pers. Soc. Psychol.* 23, 283–292. doi: 10.3389/fpsyg.2015.00250
- Efron, D. (1941/1972). *Gestures, Race And Culture*. [First edition 1941 as *Gestures and environment*. New York: King's Crown Press ed.]. The Hague: Mouton.
- Esposito, A., and Marinaro, M. (2007). “What pauses can tell us about speech and gesture partnership,” in *Fundamentals of Verbal and Nonverbal Communication and the Biometric Issue*, eds A. Esposito, M. Bratanic, E. Keller, and M. Marinaro (Amsterdam: IOS Press), 45–57.
- Fex, B., and Månsson, A.-C. (1998). The use of gestures as a compensatory strategy in adults with acquired aphasia compared to children with specific language impairment (SLI). *J. Neurolinguist.* 11, 191–206. doi: 10.1016/S0911-6044(98)00013-X
- Feyereisen, P. (1983). Manual activity during speaking in aphasic subjects. *Int. J. Psychol.* 18, 545–749. doi: 10.1080/00207598308247500
- Freedman, N. (1972). “The analysis of movement behavior during the clinical interview,” in *Studies in Dyadic Communication*, eds A. W. Siegman and B. Pope (New York, NY: Pergamon), 153–175.
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., and Wagner, S. (2001). Explaining math: gesturing lightens the load. *Psychol. Sci.* 12, 516–522. doi: 10.1111/1467-9280.00395
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.
- Goodwin, M. H., and Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica* 62, 51–75.
- Grand, S., Marcos, L. R., Freedman, N., and Barroso, F. (1977). Relation of psychopathology and bilingualism to kinesic aspects of interview behavior in schizophrenia. *J. Abnorm. Psychol.* 86, 492–500. doi: 10.1037/0021-843X.86.5.492
- Graziano, M. (2014a). “The development of two pragmatic gestures of the so-called Open Hand Supine family in Italian children,” in *From Gesture in Conversation to Visible Action as Utterance: Essays in Honor of Adam Kendon*, eds M. Seyfeddinipur and M. Gullberg (Amsterdam: Benjamin), 311–330.

- Graziano, M. (2014b). "Gestures in Southern Europe: Children's pragmatic gestures in Italy," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, eds C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and S. Tessendorf (Berlin: Mouton), 1253–1258.
- Gullberg, M. (1998). *Gesture as a Communication Strategy in Second Language Discourse: A Study of Learners of French and Swedish*. Lund: Lund University Press.
- Gullberg, M. (2006). Handling discourse: gestures, reference tracking, and communication strategies in early L2. *Lang. Learn.* 56, 155–196. doi: 10.1111/j.0023-8333.2006.00344.x
- Gullberg, M. (2011). "Multilingual multimodality: communicative difficulties and their solutions in second language use," in *Embodied Interaction: Language and Body in the Material World*, eds J. Streeck, C. Goodwin, and C. LeBaron (Cambridge: Cambridge University Press), 137–151.
- Hadar, U., Wenkert-Olenik, D., Krauss, R. M., and Soroker, N. (1998). Gestures and the processing of speech: neuropsychological evidence. *Brain Lang.* 62, 107–126. doi: 10.1006/brln.1997.1890
- Hallgren, K. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials Quant. Methods Psychol.* 8, 23–34. doi: 10.20982/tqmp.08.1.p023
- Hawkins, P. R. (1972). The syntactic location of hesitation pauses. *Lang. Speech* 14, 277–288. doi: 10.1177/002383097101400308
- Holler, J., Turner, K., and Varciana, T. (2013). It's on the tip of my fingers: co-speech gestures during lexical retrieval in different social condition. *Lang. Cogn. Process.* 28, 1509–1518. doi: 10.1080/01690965.2012.698289
- Hostetter, A. B., and Sullivan, E. L. (2011). "Gesture production during spatial tasks: it's not all about difficulty," in *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, eds L. Carlson, C. Hoelscher, and T. Shipley (Austin, TX: Cognitive Science Society), 1965–1970.
- Iverson, J. M., Capirci, O., and Caselli, M. C. (1994). From communication to language in two modalities. *Cogn. Develop.* 9, 23–43. doi: 10.1016/0885-2014(94)90018-3
- Kendon, A. (1980). "Gesture and speech: two aspects of the process of utterance," in *Nonverbal Communication and Language*, ed. M. R. Key (The Hague: Mouton), 207–227.
- Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572
- Kita, S. (2000). "How representational gestures help speaking," in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 162–185.
- Kita, S. (2009). Cross-cultural variation of speech-accompanying gesture: a review. *Lang. Cogn. Process.* 24, 145–167. doi: 10.1080/01690960802586188
- Kita, S., Alibali, M. W., and Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol. Rev.* 124, 245–266. doi: 10.1037/rev0000059
- Kita, S., Van Gijn, I., and Van der Hulst, H. (1998). "Movement phases in signs and co-speech gestures, and their transcription by human coders," in *Gesture and Sign Language in Human-Computer Interaction: International Gesture Workshop Bielefeld*, Vol. 1371, eds I. Wachsmuth and M. Fröhlich (Berlin: Springer), 23–35. doi: 10.1007/BFb0052986
- Krauss, R. (1998). Why do we gesture when we speak? *Curr. Dir. Psychol. Sci.* 7, 54–60. doi: 10.1111/1467-8721.ep13175642
- Krauss, R. K., Chen, Y., and Gottesman, R. F. (2000). "Lexical gestures and lexical access: a process model," in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 261–283.
- Krauss, R. M., and Hadar, U. (1999). "The role of speech-related arm/hand gestures in word retrieval," in *Gesture, Speech, and Sign*, eds L. S. Messing and R. Campbell (Oxford: Oxford University Press), 93–116.
- Ladewig, S. H. (2014). "The cyclic gesture," in *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*, eds C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, and S. Tessendorf (Berlin: De Gruyter Mouton), 1605–1618.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition* 14, 41–104. doi: 10.1016/0010-0277(83)90026-4
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
- Liszkowski, U. (2008). Before L1: a differentiated perspective on infant gestures. *Gesture* 8, 180–196. doi: 10.1075/gest.8.2.04lis
- MacLay, H., and Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word* 15, 19–44. doi: 10.1080/00437956.1959.11659682
- Mayberry, R., Jaques, J., and DeDe, G. (1998). What stuttering reveals about the development of the gesture-speech relationship. *New Direct. Child Dev.* 79, 77–87. doi: 10.1002/cd.23219987906
- Mayberry, R. I., and Jaques, J. (2000). "Gesture production during stuttered speech: insights into the nature of gesture-speech integration," in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 199–214.
- McNeill, D. (1985). So you think gestures are nonverbal? *Psychol. Rev.* 92, 271–295.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL: Chicago University Press.
- McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press. doi: 10.7208/chicago/9780226514642.001.0001
- Morford, M., and Goldin-Meadow, S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *J. Child Lang.* 23, 559–580. doi: 10.1017/S0305000900011569
- Morrel-Samuels, P., and Krauss, R. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *J. Exp. Psychol.* 18, 615–622. doi: 10.1037/0278-7393.18.3.615
- Morsella, E., and Krauss, R. M. (2005). Muscular activity in the arm during lexical retrieval: Implications for gesture-speech theories. *J. Psycholinguist. Res.* 34, 415–427. doi: 10.1007/s10936-005-6141-9
- Nicoladis, E., Pika, S., and Marentette, P. (2009). Do French-English bilingual children gesture more than monolingual children? *J. Psycholinguist. Res.* 38, 573–585. doi: 10.1007/s10936-009-9121-7
- Nicoladis, E., Pika, S., Yin, H., and Marentette, P. (2007). Gesture use in story recall by Chinese-English bilinguals. *Appl. Psycholinguist.* 28, 721–735. doi: 10.1017/S0142716407070385
- Özçalışkan, S., and Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition* 96, B101–B113. doi: 10.1016/j.cognition.2005.01.001
- Park-Doob, M. A. (2010). *Gesturing Through Time: Holds and Intermodal Timing in the Stream of Speech*. Available at: <https://escholarship.org/uc/item/65t9872m>
- Pine, K. J., Bird, H., and Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Develop. Sci.* 10, 747–754. doi: 10.1111/j.1467-7687.2007.00610.x
- Pizzuto, E., Capobianco, M., and Devescovi, A. (2005). Gestural-vocal deixis and representational skills in early language development. *Interact. Stud.* 6, 223–252. doi: 10.1075/is.6.2.05piz
- Poullisse, N. (1999). *Slips of the Tongue: Speech Errors in First and Second Language Production*. Amsterdam: Benjamins. doi: 10.1075/sib1.20
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ragsdale, J. D., and Silvia, C. F. (1982). Distribution of kinesic hesitation phenomena in spontaneous speech. *Lang. Speech* 25, 185–190. doi: 10.1177/002383098202500205
- Rauscher, F. H., Krauss, R. M., and Chen, Y. (1996). Gesture, speech, and lexical access: the role of lexical movements in speech production. *Psychol. Sci.* 7, 226–231. doi: 10.1002/wcs.1211
- Rose, M. (2006). The utility of arm and hand gesture in the treatment of aphasia. *Adv. Speech Lang. Pathol.* 8, 92–109. doi: 10.1080/14417040600657948
- Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language* 53, 361–382. doi: 10.1353/lan.1977.0041
- Schmid, M. S., and Fägersten, K. B. (2010). Disfluency markers in L1 attrition. *Lang. Learn.* 60, 753–791. doi: 10.1016/j.actpsy.2015.07.015
- Seyfeddinipur, M. (2006). *Disfluency: Interrupting Speech and Gesture*. Ph.D. thesis, Radboud University Nijmegen, Nijmegen.
- Seyfeddinipur, M., and Kita, S. (2001). "Gestures and disfluencies in speech," in *Oralité et Gestualité. Communication Multimodal, Interaction*, eds S. Santi, I. Guaitella, C. Cavé, and G. Konopczynski (Paris: L'Harmattan), 266–270.

- Seyfeddinipur, M., Kita, S., and Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: evidence from self-repairs. *Cognition* 108, 837–842. doi: 10.1016/j.cognition.2008.05.004
- Sikveland, R. O., and Ogden, R. (2012). Holding gestures across turns: moments to generate shared understanding. *Gesture* 12, 166–199. doi: 10.1075/gest.12.2.03sik
- Stam, G., and Tellier, M. (2017). “The sound of silence: the functions of gestures in pauses,” in *Why Gesture? How the Hands Function in Speaking, Thinking and Communicating*, eds R. Breckinridge Church, M. W. Alibali, and S. D. Kelly (Amsterdam: Benjamins), 353–377. doi: 10.1075/gs.7.17sta
- Stefanini, S., Recchia, M., and Caselli, M. C. (2008). The relationship between spontaneous gesture production and spoken lexical ability in children with Down syndrome in a naming task. *Gesture* 8, 197–218. doi: 10.1075/gest.8.2.05ste
- The Pygos Group (1992). *Pingu's Family Celebrates Christmas*.
- Trofimovich, P., and Baker, W. (2006). Learning second language suprasegmentals: Effects of L2 experience on prosody and fluency characteristics of L2 speech. *Stud. Sec. Lang. Acquisit.* 28, 1–30. doi: 10.1017/S0272263106060013
- Volterra, V., Bates, E., Benigni, L., Bretherton, I., and Camaioni, L. (1979). “First words in language and action: a qualitative look,” in *The Emergence of Symbols: Cognition and Communication in Infancy*, ed. E. Bates (New York, NY: Academic Press), 141–222.
- Willems, R. M., and Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: a review. *Brain Lang.* 101, 278–289. doi: 10.1016/j.bandl.2007.03.004
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). “ELAN: a professional framework for multimodality research,” in *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, Genoa.
- Yasinnik, Y., Shattuck-Hufnagel, S., and Veilleux, N. (2005). *Gesture Marking of Disfluencies in Spontaneous Speech*. Available at: http://www.isca-speech.org/archive_open/archive_papers/diss_05/dis5_173.pdf

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Graziano and Gullberg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership