

DE GRUYTER

Désirée Kleineberg

THE EXPRESSION OF "COLLECTIVITY" IN ROMANCE LANGUAGES

AN EMPIRICAL ANALYSIS OF NOMINAL
ASPECTUALITY WITH FOCUS ON FRENCH

BEIHEFTE ZUR ZEITSCHRIFT FÜR
ROMANISCHE PHILOLOGIE EDITION NIEMEYER

Désirée Kleineberg

The expression of “collectivity” in Romance languages

Beihefte zur Zeitschrift für romanische Philologie



Herausgegeben von
Éva Buchi, Claudia Polzin-Haumann, Elton Prifti
und Wolfgang Schweickard

Band 472

Désirée Kleineberg

The expression of “collectivity” in Romance languages

An empirical analysis of nominal aspectuality
with focus on French

DE GRUYTER

Zgl. phil. Diss. Universität Tübingen 2021

This open access publication has received funding from the SFB 833 “The construction of meaning” (Tübingen University, Germany).

ISBN 978-3-11-078458-9

e-ISBN (PDF) 978-3-11-078469-5

e-ISBN (EPUB) 978-3-11-078479-4

ISSN 0084-5396

DOI <https://doi.org/10.1515/9783110784695>



This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. For details go to <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Library of Congress Control Number: 2022936739

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2022 Désirée Kleineberg, published by Walter de Gruyter GmbH, Berlin/Boston

The book is published open access at www.degruyter.com.

Typesetting: Integra Software Services Pvt. Ltd.

Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Acknowledgements

This book represents a slightly revised version of my PhD thesis which was successfully submitted and defended at the Faculty of Humanities of the University of Tübingen in 2021. Although doing a doctorate is often associated with working alone in an ivory tower, this dissertation was facilitated and even made possible by a number of people. I have done my PhD in the framework of the project “Verbal and nominal aspectuality between lexicon and grammar” as part of the Collaborative Research Centre 833 at the University of Tübingen (Project-ID 75650358 – SFB 833). First and foremost I am very grateful to the *Deutsche Forschungsgemeinschaft* for its generous financial support. With this, not only the majority of experiments which form the empirical basis of this dissertation but also some very fruitful trips to conferences and summer schools, which enabled me to discuss my results and broaden my scientific horizon, were made possible.

I also want to heartily thank my first supervisor Wiltrud Mihatsch who accompanied my studies of Romance linguistics from my first lectures at the University of Bochum until today. She not only inspired my interest in linguistics, she always challenged me and at the same time removed every possible obstacle in my path. I am also very grateful to my second supervisor Sarah Dessì Schmid for the constant encouragement and the fruitful discussions. I am very grateful to both of them for being such encouraging, fastidious, warm and helpful supervisors. A very warm thanks also to Lydia Momma who accompanied me during the whole PhD-process not only as a colleague, but as a friend.

Furthermore, I want to thank Éva Buchi, Claudia Polzin-Haumann, Elton Prifti and Wolfgang Schweickard for accepting the book for their series, Ulrike Krauß and Christine Henschel of De Gruyter for their helpful support during the publication process, as well as Scott Martingell for doing the English proofreading.

A warm thanks also to all my colleagues of the Romance department of the University of Tübingen as well as of the Collaborative Research Centre 833. Thank you for the fruitful discussions and helpful pieces of advice in the colloquiums, during lunch and in hallway conversations. My biggest thanks also to my friends and family for their encouragement, for supporting my moods and my “if you want to know it exactly...” lectures. Last, but absolutely not least, I want to thank my dear husband Sven. For over ten years now he never left my side, celebrated good times with me and supported me in moments of doubt. Without him, this book would probably never would have been completed.

Werl, March 2022

Contents

Acknowledgements — V

List of figures — XI

List of tables — XV

List of abbreviations — XVII

Introduction — 1

I Theoretical foundations

1 Nominal aspectuality and number — 11

1.1 The construal of extra-linguistic entities — 12

1.2 The linguistic expression of nominal aspectuality — 22

1.2.1 The typological perspective — 23

1.2.2 Nominal aspectuality in Romance languages — 28

2 Interim summary: From nominal aspectuality to collection nouns — 35

3 Collection nouns: State of the art and definitional delimitation — 37

3.1 Collection nouns as a semantic-syntactic category — 37

3.1.1 External boundedness of count collective nouns and object mass nouns — 40

3.1.2 Internal plurality of count collective nouns and object mass nouns — 45

3.1.3 An integral semantic-syntactic model of collections — 61

3.2 Collection nouns as a morphological category — 66

II The synchronic characteristics of collection nouns in present-day language

4 Semantic-syntactic characteristics: An acceptability judgement study on count collective nouns and object mass nouns — 85

4.1 Methodology — 86

4.2 Results — 104

- 4.2.1 Testing the external boundedness of count collective nouns and object mass nouns — **104**
- 4.2.2 Testing the internal plurality of count collective nouns and object mass nouns — **120**
- 4.2.3 Revising the integral semantic-syntactic model of collections — **141**

5 Morphological characteristics: A corpus analysis of collective nonce-formations — 147

- 5.1 Theoretical preliminaries: Construction morphology — **150**
- 5.2 Methodology — **160**
- 5.3 Results — **164**
 - 5.3.1 French: Collective nonce-formations in *frTenTen12* — **167**
 - 5.3.2 Spanish: Collective nonce-formations in *esTenTen11* — **172**
 - 5.3.3 Italian: Collective nonce-formations in *itTenTen16* — **180**
 - 5.3.4 Pan-Romance comparison — **187**

6 Conclusion: Linguistic characterisation of Romance collection nouns in present-day language — 197

III The diachronic development of French collection nouns

7 State of the art: The assumed paths of lexicalisation of collection nouns — 205

- 7.1 The concept of lexicalisation — **205**
- 7.2 The phylogeny of collection nouns — **210**
 - 7.2.1 Traditional historical linguistics — **211**
 - 7.2.2 Cognitive historical linguistics — **216**
 - 7.2.3 Artefacts as SOMNs — **220**
- 7.3 The ontogeny of collection nouns — **221**

8 Lexicalisation of collection nouns: Corpus analysis in *Frantext* — 227

- 8.1 Methodology — **227**
- 8.2 The etymology of collection nouns — **240**
- 8.3 Results of the corpus analysis — **243**
 - 8.3.1 Event-related artefact collection nouns: *habillement, accoutrement, vêtement, équipement, mercerie, fringues* — **243**

- 8.3.2 Collection nouns based on one shared property: *mobilier*,
gens — 265
- 8.3.3 Collection nouns based on the Latin neuter plural: *vaisselle*, *bétail*,
volaille — 276

- 9 **Conclusion and discussion: Linguistic characterisation of collection
nouns in language evolution — 285**

IV General conclusion and outlook

- 10 **Summary of the results and conclusion — 289**

- 11 **Outlook — 295**

- 12 **References — 297**
 - 12.1 Corpora — 297
 - 12.2 Dictionaries and encyclopaedias — 297
 - 12.3 Research literature — 298

- Subject index — 317**

List of figures

- Figure 3.1** Schematic representation of different types of collection noun — 64
- Figure 4.1** Example rating in the acceptability judgement study — 99
- Figure 4.2** Mean acceptability ratings of Romance OMNs in combination with the indefinite article — 107
- Figure 4.3** Mean acceptability ratings of the construction <indef. art. + OMN> in the four tested Romance languages in the follow-up study — 110
- Figure 4.4** Mean acceptability ratings of the construction <neutral verb + indef. art. + OMN> in the four tested Romance languages in the follow-up study — 112
- Figure 4.5** Mean acceptability ratings of the construction <non-distributive quantifier + SOMN_{pl}> in Italian, Spanish and Portuguese — 116
- Figure 4.6** Mean acceptability ratings of French POMNs in combination with small round numbers vs. *plusieurs* 'various' — 118
- Figure 4.7** Mean acceptability rating of French and Italian collection nouns in combination with a stubbornly distributive predicate — 122
- Figure 4.8** Mean acceptability rating of French and Italian collection nouns in combination with a stubbornly distributive predicate in the follow-up-study — 124
- Figure 4.9** Mean acceptability rating of inanimate Romance OMNs and CCNs in combination with highly distributive predicates — 128
- Figure 4.10** Mean acceptability rating of animate Romance OMNs and CCNs in combination with highly distributive predicates — 130
- Figure 4.11** Mean acceptability rating of human Romance OMNs and CCNs in combination with highly distributive predicates — 131
- Figure 4.12** Mean acceptability rating of Romance SOMNs and CCNs in combination with highly distributive predicates — 133
- Figure 4.13** Mean acceptability rating of Romance 'furniture'-denoting SOMNs with highly distributive predicates — 136
- Figure 4.14** Mean acceptability rating of Romance OMNs in hyponymic and meronymic constructions — 140
- Figure 4.15** Revised integral semantic-syntactic model of Romance collection nouns — 145
- Figure 5.1** Schematisation of English nouns on *-er* (Booij 2016, 430) — 154
- Figure 5.2** First draft of a constructional schema <collection> in French — 156
- Figure 5.3** Semantic map of typical dative functions/functional scope of Engl. *to* and Fr. *à* (adapt. from Haspelmath 2003, 213–215) — 158
- Figure 5.4** Example of the corpus configurations in *frTenTen12* for the suffix *-ment* — 162
- Figure 5.5** Semantics of unique nonce-formations in *frTenTen12* (1,000-word sample extrapolated to all hapaxes) — 168
- Figure 5.6** Semantics of unique nonce-formations in *esTenTen11* (1,000-word sample extrapolated to all hapaxes) — 173
- Figure 5.7** Semantics of unique nonce-formations in *itTenTen16* (1,000-word sample extrapolated to all hapaxes) — 180
- Figure 5.8** Distribution of mass and count determiners in combination with collective ad hoc constructs — 190
- Figure 5.9** Semantic map of collectivity in Romance — 194

- Figure 5.10** Semantic map of the functional range of collective suffixes in Romance — **195**
- Figure 7.1** Lexicalisation paths of derived collection nouns in Romance — **215**
- Figure 7.2** Assumed pathway of lexicalisation of collections (Mihatsch 2016, 301) — **219**
- Figure 8.1** Evolution of number inflection of Fr. *habillement* in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **244**
- Figure 8.2** Evolution of types of determiner in combination with Fr. *habillement*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **247**
- Figure 8.3** Evolution of nominal aspectual types of Fr. *habillement*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **248**
- Figure 8.4** Evolution of the nominal types of Fr. *mercerie*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **251**
- Figure 8.5** Evolution of number inflection of Fr. *accoutrement* in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **253**
- Figure 8.6** Evolution of types of determiner in combination with Fr. *accoutrement*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **253**
- Figure 8.7** Evolution of the nominal types of Fr. *accoutrement*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **255**
- Figure 8.8** Evolution of the nominal types of Fr. *équipement*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **257**
- Figure 8.9** Evolution of number inflection and of the nominal types of Fr. *fringues* in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **259**
- Figure 8.10** Evolution of the nominal types of Fr. *vêtement*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **262**
- Figure 8.11** Evolution of types of determiner in combination with Fr. *vêtement*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **262**
- Figure 8.12** Evolution of number inflection of Fr. *vêtement* in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **263**
- Figure 8.13** Evolution of number inflection of Fr. *meuble* in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **267**
- Figure 8.14** Evolution of types of determiner in combination with Fr. *meuble*_{SG} and evolution of the nominal types of Fr. *meuble*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **270**
- Figure 8.15** Evolution of number inflection of Fr. *gens* in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **271**
- Figure 8.16** Evolution of types of determiner in combination with Fr. *gent*_{SG} and evolution of the nominal types of Fr. *gent*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **273**
- Figure 8.17** Evolution of types of determiner in combination with Fr. *gens*_{PL} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **274**
- Figure 8.18** Evolution of number inflection of Fr. *vaisselle* in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **277**
- Figure 8.19** Evolution of types of determiner in combination with Fr. *vaisselle*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **278**
- Figure 8.20** Evolution of number inflection of Fr. *bétail* in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **280**

- Figure 8.21** Evolution of types of determiner in combination with Fr. *bétail*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **281**
- Figure 8.22** Evolution of number inflection of Fr. *volaille* in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **282**
- Figure 8.23** Evolution of types of determiner in combination with Fr. *volaille*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000) — **283**

List of tables

Table 1.1	Aktionsarten according to Vendler (1957) (table adapt. from Boogaart 2004, 1169) — 17
Table 1.2	The coding of the mass-count distinction in various Romance languages — 34
Table 3.1	Romance collective suffixes and their semantic nuances — 71
Table 4.1	Comparison of token quantities and text types of different Romance corpora focusing on present-day language (token quantities are rounded) — 88
Table 4.2	List of potential Romance OMNs (SG/PL – percentage of occurrences with the indef. art.) — 91
Table 4.3	Romance OMNs and their quasi-synonyms in the acceptability judgement study — 94
Table 4.4	Overview of the participants of the acceptability judgement study — 99
Table 4.5	Overview of the participants of the follow-up acceptability judgement study — 103
Table 5.1	Quantitative overview of the ad hoc derivatives based on potentially collective suffixes in the <i>TenTen</i> corpora — 165
Table 8.1	Frequencies of the collection nouns in <i>Frantext</i> and sample sizes analysed in the diachronic study — 232
Table 8.2	Classification of determiners and quantifiers in the diachronic analysis — 238

List of abbreviations

BrPt.	Brazilian Portuguese
CCN	count collective noun
Engl.	English
EuPt.	European Portuguese
Fr.	French
Germ.	German
It.	Italian
Lat.	Latin
OMN	object mass noun
POMN	plural object mass noun
PON	plural object noun
SOMN	singular object mass noun
SON	singular object noun
Sp.	Spanish

Introduction

Languages typically have different inventories of various kinds of word by means of which the multifaceted domains of our human experiences can be verbalised. The study of and first approaches to categorising these parts of speech go back to classical Greek and Latin grammarians like Dionysios Thrax or Varro. They mainly based their categorisation of nouns, verbs and other word classes on semantic, syntactic and morphological criteria and had considerable impact on all linguistic descriptions following them. However, by the time linguists became more familiar with structurally different languages like the American Indian languages, it became clear that the classical terminology and methodology of classification was less useful for languages other than Greek, Latin, their descendants and linguistic relatives (cf. Broschart 2002, 663). Abstracting from language-specific morpho-syntactic characteristics, modern typologists nowadays tend rather to base their categorisation of word class membership on conceptual-semantic features. Adopting this perspective, one may generalise that most, if not all, languages of the world have at least some linguistic means of expression of reference to objects (a noun class), of predication of processes (a verb class), and of attributing qualities (an adjective class), whatever the formal reflexion of these categories may look like (cf. Broschart 2002, 666). More specifically, “grammatical behavior [. . .] is best regarded as SYMPTOMATIC of its semantic value, not the sole or final basis for a criterial definition” (Langacker 1991, 60; original stress). It follows from this line of argument that there are representatives of a word class like that of nouns which are more or less prototypical. A typical noun denotes physical objects, i.e. time-stable, concrete and coherent individuals with a reduced number of features, which are non-relational (cf. Langacker 1991, chap. 3; Croft 2001, 86–92; Givón 2018, 249–250) and are thus situated at the basic level of categorisation, i.e. “the level at which categories carry the most information, possess the highest cue validity, and are, thus, the most differentiated from one another” (Rosch et al. 1976, 383). Evidence for the prototypicality of nouns like *apple* or *chair* comes from various studies showing that they are e.g. acquired earlier (cf. Bloom 2000, 91–92; Gentner/Boroditsky 2001) and are used predominantly for object naming, as opposed to other nouns at a higher or lower level of generalisation (cf. Rosch et al. 1976, 423–424). In contrast, atypical nouns like *equipment*, *truth*, *sand* and *people* denote referents that are less spatially coherent and more complex than single individuals and that do not represent good gestalts in not necessarily being holistically perceivable nor standing out from the ground as figures. Within this category of atypical nouns, there is a group whose conceptual complexity lies in their denoting a plurality

of more or less coherent entities. These kinds of noun may vary both inter- and intra-linguistically in their specific formal characteristics (cf. Mihatsch 2015a for an overview), but what they all have in common is the semantic feature [+ plural] verbalised in one single lexical entry. It must be specified that it is not only plurality that they express but, more specifically, these kinds of noun express the concept of COLLECTIVITY. In other words, a noun like *people* does not just denote a plurality of persons, but a plurality of persons sharing some feature, be it time, space, a property or function (cf. Meisterfeld 1998, 50–51). For this reason I call these nouns, following Kuhn (1982) and Seiler (1986), *collection nouns*. There are various types of collection noun in the Indo-European languages which differ mainly in their conceptualisation of the plurality of referents as well as in the morphological form the plurality is verbalised with. Firstly, there are countable collective nouns (henceforth CCNs) like Sp. *atuendo* ‘outfit’, Engl. *team* or Fr. *essaim* ‘swarm’. Those nouns refer in the inflectional singular to a set of things, but the plural form does not denote more of those things, but more of those collections. Semantically, these kinds of collection are bounded groups, of which the referents form a part (cf. Flaux 1999). Secondly, there are singular object mass nouns (henceforth SOMNs) like Sp. *ropa* ‘clothing’, Germ. *Obst* ‘fruit’ or Engl. *furniture*, that syntactically behave like homogeneous mass nouns such as *water* – they are uncountable and incompatible with the indefinite article – but semantically they too refer to a plurality of artefacts, animals or human beings. These kinds of collection are not holistically perceivable groups, but rather kind-denoting pluralities (cf. Rothstein 2010a). SOMNs only appear in the morphological singular with no equivalent plural form, but we do also find plural object mass nouns (henceforth POMNs) like Engl. *groceries* or Germ. *Leute* ‘people’ that have a fossilised plural form, but which can hardly be counted and have no equivalent singular form (cf. Acquaviva 2008; Lauwers 2014).

The two Spanish examples of *atuendo* ‘outfit’ and *ropa* ‘clothing’ and also their English equivalents show that a certain language may encode the same extra-linguistic referents – like t-shirt, trousers and shoes – differently, either as a clearly delimited or as a vaguer and more substance-like collection. This kind of linguistic construal is typically described by the concept of *apprehension* (cf. Kuhn 1982; Seiler 1986) or *nominal aspect* (cf. Meisterfeld 1998). Seiler (1986, 9) defines *apprehension* as “[. . .] the universal operational dimension with corresponding subdimensions which explains how language grasps and represents concepts that correspond to objects or items.”

In a similar way, and basing himself on research on nominal number (e.g. Biermann 1982, but also Damourette/Pichon 1911–1927), Meisterfeld (1998, 34) defines the notion of *Nominalaspekt* as the grammatical reflection of a speaker’s perspective or view on an extralinguistic object. These definitions thus pattern

with the understanding of *construal* as it is defined in cognitive linguistics: “[. . .] [the] ability to conceive and portray the same situation in alternate ways” (Langacker 2008, 43). Meisterfeld (1998) specifies this notion of nominal aspect in differentiating between an inner and an outer nominal aspect. The former refers to the internal structure of an entity (cf. Meisterfeld 1998, 44), the latter to its external boundaries (cf. Meisterfeld 1998, 36; cf. Jackendoff 1991a; Rijkhoff 2002 for a similar distinction; their similarities and differences will be discussed in chap. 1.1). By means of this distinction the differences between the various noun types can be described, but it also legitimises their subsumption under one umbrella term, *collection nouns*: With respect to their inner aspect, all collection nouns are amalgams – that is, they denote a plurality of discrete entities that in some way cohere more than is implied by the mere inflectional plural. As Gil (1996, 63) puts it: “[. . .], even though *the boys* may be coextensive with *the team*, the latter NPs says more, namely that the boys are organized in a particular fashion.” Gil (1996) distinguishes in this respect between an additive plural (*boys*) and a non-additive plural (*team*), whereby *team* is non-additive since it is more than the mere sum of its parts. Concerning the outer nominal aspect, collection nouns differ in whether they designate bounded or unbounded collections, or *Gruppen-* vs. *Genuskollektiva* (‘group collectives’ vs. ‘type collectives’, in the terminology of Leisi 1975). Following Kuhn (1982) and Seiler (1986), I choose the notion of collection which the latter defines as a “technique [which] is based on the relationality between the individual (set) and a unity of individuals characterized by certain Gestalt qualities” (Seiler 1986, 59). Generalising slightly and connecting the theories of Seiler and Meisterfeld, I define a collection as a coherent plurality of entities that is either construed as bounded with certain gestalt qualities or as an unbounded mass. Collections are mostly denoted by collection nouns. This approach will be explained in more detail in the following chapters.

CCNs have long been at the centre of research on French, English and to a lesser degree other Romance languages like Spanish and Catalan, where they have been examined from different perspectives and in the framework of various linguistic traditions (cf. i.a. Quirk 1985; Michaux 1992; Bosque 1999; Flaux 1999; Levin 2001; Solé Solé 2002; Lammert 2010). SOMNs, however, have only recently become the focus of linguistic research, particularly in the framework of formal semantics and concentrating mainly on languages like English or Mandarin Chinese, but scarcely on Romance languages (Chierchia 2010; Rothstein 2010a; Landman 2011). Since the seminal work of Acquaviva (2008) on lexical plurals there has been growing interest in this nominal type culminating in a recent special issue of *Linguisticæ Investigationes on Lexical plurals and beyond* (Lauwers/Lammert 2016). Lexical plurals are understood as intrinsically plural, i.e. grammatically and semantically (cf. Acquaviva 2008, 268). In contrast to the inflectional plural, lexical plurality

is inherent to a noun in not being determined by the syntactic context, but by the semantic choice of a speaker in a certain context. Lexical plurality is, thus, not obligatory, nor general, in the sense that it does not apply to all nouns and pronouns. Examples are for instance Engl. *brains* ‘intelligence’, Dutch *letteren* ‘literature’ or It. *mura* ‘walled perimeter’. These nouns have in common that they not only are morphologically plural, but also that they have a default semantic plurality. They often, but not necessarily, co-exist with an inflectional plural form, which also differs in meaning like It. *muri* ‘walls’. POMNs are thus analysed in this framework as a subcategory of lexical plurals that refer to clearly distinguishable entities (cf. Mihatsch 2016) and which are consequently distinct from other kinds of inherent plurality like nouns denoting granular aggregates.

What these analyses, whether recent or more traditional, have in common is that they focus mainly on only one type of collection noun and on only one language; cross-linguistic comparison has so far not been at the centre of interest – comparative research as found in Joosten (2006) or Mihatsch (2016), for instance, is therefore hard to find. For Romance languages, it is particularly the French language that has been analysed, while collection nouns in other languages (like Spanish, Italian or Portuguese) have only been mentioned rather briefly in general overviews or in the framework of related research areas (cf. i.a. Ortega/Morera 1981–1982; Bosque 1999; Grossmann 2004). Solé Solé (2002) presents an integral approach on Catalan CCNs. Additionally, what have been mainly focused on in Romance philology are bounded CCNs, OMNs have not in general been treated (but see the recent work of Lauwers 2014). Furthermore, the majority of works examining collection nouns focuses on present-day language use and the nouns’ syntactic and semantic characteristics. The diachronic evolution of collection nouns has so far only been outlined on the basis of a few qualitative corpus analyses and etymological dictionaries (cf. Mihatsch 2006; 2016; Grimm/Levin 2011; 2012; 2016). More elaborate examinations like the ones of Baldinger (1950) or Collin (1918) are – given their age – not necessarily outdated but by and large limited in their empirical possibilities. The focus on the semantic-syntactic aspects of collection nouns finally does not take them into account as a possible morphological category. As a consequence of these considerations, several research gaps arise which I will address in this present study:

- There is barely any research considering collection nouns as a category with related sub-types, i.e. CCNs, SOMNs and POMNs. Research done to date either mostly focused on one of these sub-types, neglecting the other, or considered particularly the category of OMNs as a quirky sub-type of CCNs. Additionally, whereas CCNs have already been analysed quite exhaustively (for French), there has been nearly no research undertaken on OMNs in Romance languages.

- There are hardly any cross-linguistic comparisons of collection nouns, not even for one of the nominal sub-types. Comparative work has mostly only focused on a rough comparison between English and French.
- There has been little research done on morphological aspects of collection nouns (cf. Mihatsch 2021); research done so far has mostly concentrated on just the semantic and syntactic characteristics.
- There is a scarcity of empirical research on the diachronic development of collection nouns, with the research done so far being mainly theoretical.

Following on from these research gaps, the present work has two main goals which can be mapped onto two fields. On the one hand, I will give an overview as complete as possible of the linguistic characteristics of Romance collection nouns in the synchrony of contemporary language. This includes a comparison of different types of collection noun in one language, as well as their cross-linguistic comparison in this language family. I will furthermore address the question of collection nouns as a possible derivational category. On the other hand, I will examine the diachronic development of Romance collection nouns, placing the theoretical findings on an empirical basis. There are thus three central research questions:

RQ1: What are the influencing factors on the particular linguistic expression of a collection of entities and the semantic-syntactic characteristics related to it?

RQ2: To what extent are there any productive word-formation patterns in the domain of COLLECTIVITY?

RQ3: To what extent do collection nouns follow a unidirectional path of lexicalisation? Can this path be empirically proven?

Since the main goal of this work lies mainly in the bringing together the various theoretical accounts and not in focusing only on one topic, a cross-linguistic comparison between several Romance languages is unfortunately not possible in every research domain that will be examined. Instead, I will take French as a focal point and compare it to other Romance languages whenever this is possible and fruitful. The concentration on French is done for two reasons: firstly, French is a language with a fully grammaticalised distinction between mass and count syntax (for example by the partitive article unambiguously marking mass nouns). This feature allows for a clear definitional delimitation between various types of collection noun not only on the basis of their semantic characteristics (e.g. bounded groups vs. indefinite numbers of referents) but also on syntactic grounds. This adds another possible piece to the puzzle unavailable to that extent in languages like Spanish or Portuguese that do not use such a mass-marking

determiner, or Italian where the partitive article displays a lesser degree of obligatoriness (cf. Stark 2008 for an overview). The establishment of clearly delimited categories of collection noun in French may thus serve well as a starting point for additional cross-linguistic comparisons. Secondly, as mentioned above, research on collection nouns in Romance languages has mainly focused on French. At first sight, it seems to be counterintuitive to concentrate on a language that has already been quite extensively examined. However, this predominantly theory-orientated research has never been empirically tested and consolidated – in the sense of corpus-based approaches and studies involving more than one (non-expert) speaker. By focusing on French, it is possible to develop a complete theory regarding collection nouns that has been empirically confirmed and which may then serve as a bridge to other Romance languages having a well consolidated theory as a basis of argument.

This monograph is divided into three major sections: Part I will lay the theoretical groundwork for the following parts. I will explain in more detail the aspects mentioned above, focusing on nominal aspectuality and number (chap. 1), the relation between nominal aspectuality and collection nouns (chap. 2) and a definitional delimitation of the category *collection noun* based on the state of current research (chap. 3). This overview of the state of the art will focus both on semantic-syntactic issues as well as those concentrating on word-formation. The section will not only be limited to Romance languages but will generally adopt a typological perspective. The second part will treat collection nouns in present-day language use in Romance languages. Here, I will focus on two major empirical domains: I will first analyse the semantic and syntactic characteristics of CCNs and OMNs in French, comparing them to equivalent nouns in Spanish, Italian and Portuguese, focusing thus mainly on the comparative aspect of analysis. The empirical basis here will be an acceptability judgement study (chap. 4). Second, I will examine Romance collection nouns from a morphological point of view, analysing collective nonce-formations in French, in comparison to Spanish and Italian (chap. 5). The aim of part II is to develop a complete model of the concept of COLLECTIVITY and its linguistic means of expression in contemporary use in Romance languages (presented in chap. 6). The exact choice of languages compared to French will be explained in detail in the respective sections. Part III will finally address the question of the diachronic development of Romance collection nouns. Here, I will mainly analyse the theory of a unidirectional lexicalisation path of collection nouns, developed by Mihatsch (2006; 2016). After discussing the concept of lexicalisation and presenting the state of the art on the assumed pathways of evolution of collection nouns (chap. 7), I will concentrate on a corpus analysis of this nominal type in French to investigate theoretical assumptions regarding this possible lexicalisation path (chap. 8). The empirical studies

presented in parts II and III were conducted in the framework of and financed by the research project “Verbal and nominal aspectuality between lexicon and grammar” (directed by Prof. Sarah Dessì Schmid and Prof. Wiltrud Mihatsch) as part of the Collaborative Research Centre 833 at the University of Tübingen, funded by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – SFB 833 – Project-ID 75650358.

The present work will conclude by presenting a model of COLLECTIVITY in Romance languages that takes into consideration several linguistic domains, cross-linguistic variation and diachronic evolution. COLLECTIVITY in Romance languages will be shown to be governed by the principle of continuity: continuity between different types of collection, continuity between different Romance languages and continuity in historical development.



I Theoretical foundations

1 Nominal aspectuality and number

As has been already indicated in the introduction, the concept of nominal aspectuality will provide the theoretical foundation to describe and to analyse the linguistic means of expression of COLLECTIVITY in Romance languages. In what follows, I will thus first discuss different approaches to the construal of extra-linguistic entities to then develop my own theory of nominal aspectuality as it will be understood in this framework (chap. 1.1). Chap. 1.2 will then discuss the different means of expression of nominal aspectuality, not only in the Romance languages, but also from a typological point of view. A crucial aspect in this description will be the marking of nominal number. As has been mentioned above, collection nouns differ, among other aspects, in verbalising a collection either as a bounded group or an indefinite plurality. We will see shortly that this constitution of the outer nominal aspect correlates strongly with the compatibility with certain morpho-syntactic operations implying countability, thus i.a. nominal number. This already indicates a very important theoretical premise: As will be shown in chap. 1.1, I adopt an onomasiological conception of nominal aspectuality in analysing the different linguistic means of expressing COLLECTIVITY. This implies in turn the assumption of three different layers of analysis: first, there is the conceptual layer, supposedly independent of language. Second, there are cross-linguistically more or less universal linguistic categories to express this conceptual layer, like e.g. nouns or articles. And third, there are the language-specific means of expression which underlie diasystematic variation or certain syntactic restrictions. These three layers are, however, often difficult to separate neatly since it is hard to describe the conceptual layer completely without any linguistic means. Following Koch (2003), I thus opt for the adoption of the perspective of an *onomasiologie éclairée* ('enlightened onomasiology'). Consequently, I will describe in what follows my understanding of this onomasiological category of nominal aspectuality, which cannot just simply be described without instrumentalising the linguistic means verbalising it. In a first step, I will thus refer to various illustrating and mainly English examples (cf. chap. 1.1). These elaborations will then be systematised in considering not only the group of Romance languages (cf. chap. 1.2.2), but also the typological perspective (cf. chap. 1.2.1). With this, the theoretical conception of the onomasiological category may then be controlled and, if necessary, revised: "Durch den interlingualen Vergleich vermeidet man es, ein bestimmtes einzelsprachlich vorfindliches Netz versprachlichter Konzepte als universal zu setzen" ('the cross-linguistic comparison prevents the considering of a certain language-specific network of verbalised concepts as universal') (Blank/Koch 2003a, 7).

1.1 The construal of extra-linguistic entities

Before coming to the theoretical elaboration of nominal aspectuality, some brief remarks on the methodological and theoretical framework are necessary. As indicated by the examples given in the introduction, the present study adopts a cognitive linguistic perspective on the classification of various nominal types, i.e. individual count nouns, collection nouns, substance denoting nouns etc. There is, however, an alternative approach to the classification of noun types represented by numerous studies coming from formal semantics and focusing mainly on the opposition between mass and count nouns (cf. i.a. Bunt 1979; Krifka 1989; Chierchia 1998a; 2010; Rothstein 2010a; Landman 2011). Both theoretical perspectives make, as we will shortly see, the same observations on language. One very prominent observation is e.g. that the same extra-linguistic entities may be labelled by different quasi-synonyms, more specifically by a count noun and a mass noun. This is not only true for quasi-synonymous pairs in one language (e.g. *shoes* vs. *footwear*), but also between languages (e.g. Engl. *furniture*_{mass} vs. It. *mobile*_{count}) (cf. e.g. Chierchia 1998a, 56 for formal semantics and cf. e.g. Croft/Cruise 2004, 41 for cognitive semantics). The difference between these two theoretical approaches lies in the lines of argument following from these observations. While cognitive semanticists opt for an explanation of these regularities by taking into consideration more general principles of thinking and speaking, formal semanticists focus on the formal description of the underlying linguistic structures of different nominal types. In the following, I adopt the former perspective since the main goal of the present study lies in analysing nominal types using a holistic approach, not only focusing on the syntax-semantics interface, but also on derivational aspects, usage frequency, connotation and diatopic as well as diachronic variation. Formal semantic studies will be consulted as important sources for various theoretical domains, particularly in the framework of the description of the semantic-syntactic characteristics of collection nouns in present-day language use (chap. 3.1), but the further theory-building will be based on the assumptions of cognitive linguistics.

Adopting this cognitive approach, the quasi-synonymous pairs mentioned above are explained by different ways of construing the same extra-linguistic entities. On a semasiological level, expressions like Sp. *atuendo* ‘outfit’ and Sp. *ropa* ‘clothing’ are thus said to reflect the way of interpreting the referents (cf. Croft/Cruise 2004, 3–4; Langacker 2008; Ising 2019). This is not only true for the nominal but also for the verbal domain (examples taken from Croft 1998, 69; Croft/Cruise 2004, 41):

- (1) a. leaves on the tree
- b. foliage-∅ on the tree

- (2) a. Tess is playing the flute.
 b. Tess plays the flute.

Examples (1) and (2) exemplify what Talmy (1988) calls *structural schematization* which “comprises all the forms of conceptual delineation that can be ascribed to a quantity, or to the pattern in which two or more quantities are interrelated, whether in space or time or some other conceptual dimension” (Talmy 1988, 194). Structural schematisation consequently means that, by using the basically truth-conditionally equivalent but inflectionally different expressions in (1a–b) and (2a–b), we simply construe the situations and objects differently. By using the count noun *leaf* in (1a) we construe the extra-linguistic objects as individual bounded entities, and by using the mass noun *foliage* in (1b) we, in contrast, interpret the same extra-linguistic entities rather as a homogeneous substance (cf. Croft/Cruse 2004, 64). The same holds for the introductory example of Sp. *atuendo* and *ropa*: a t-shirt, trousers and a pair of shoes can be either construed as a complete set of clothes or as a theoretically open collection, which may in turn be reflected by the choice of different linguistic means of expression. In example (2a) it is the momentarily executed action of playing the flute which is focused on, whereas example (2b) construes this same action as habitual – the basic action of playing the flute, however, is the same in both examples. Langacker (2008, 131) summarises this: “We are perfectly capable of construing the same conceived entity in alternate ways, each of which highlights certain aspects of it and downplays others.”

This construal of the extra-linguistic reality surrounding us can be described by numerous theories that vary in their theoretical orientations. First of all, there are the already mentioned cognitive approaches of Talmy (1988), Croft/Cruse (2004), Langacker (2008), Ising (2019), but also Jackendoff (1991a). These assume that language construes the extra-linguistic world and does not directly reflect it. Related to these, but concentrating more on typology and language-specific properties, are the theories of Hansjakob Seiler and the Cologne UNITYP project, Meisterfeld (1998), as well as Rijkhoff (1991; 2002), which all focus on the question of how the world’s languages verbalise objects, persons and events. All these approaches, which I will address in more detail later, have in common that they assume some opposition between countable and uncountable nouns which reflect the construal of a named entity either as bounded or as unbounded (cf. *atuendo* vs. *ropa*). This distinction has already been made by Bloomfield (1933, 205), Damourette/Pichon (1911–1927) and Jespersen (1949), so it is neither new nor exclusively characteristic of cognitive linguistics. In addition to the dichotomous opposition of boundedness, the majority of authors also mention some kind of internal configuration, mostly described in terms of homogeneity or

heterogeneity. In this respect, nouns can be characterised as either interpreting the referent as homogeneous like *water* or *boy*, or as being constituted of heterogeneous entities like *team* or *cattle* (cf. Langacker 1987b; 2008, 139–142; Jackendoff 1991a, 18–20; Meisterfeld 1998, 36; Rijkhoff 2002, 50–56). These terms are best described by the distinction between homogeneous referents, whose constituting entities are not distinguishable or not of concern, and heterogeneous referents, whose constituting entities are conceptually prominent because of their e.g. functional or perceptual importance. In other words, substances and single individuals are not construed as being constituted of distinguishable parts since these constituents are either too small (*water*) or not of concern since the referent is perceived holistically (*boy*). In contrast, pluralities like *team* or *cattle* are construed as having an internal structure of distinguishable entities. The two basic dichotomies – boundedness and internal structure – are generally combined to classify four conceptual entity types that correlate with different noun types, here exemplified by the typology of Jackendoff (1991a, 20):

- (3) +bounded, –internal structure: Individuals (*a pig*)
 +bounded, +internal structure: Groups (*a committee*)
 –bounded, –internal structure: Substances (*water*)
 –bounded, +internal structure: Aggregates (*buses, cattle*)

Collections verbalised by collection nouns are to be classified as having an internal structure, thus as being heterogeneous. To be able to distinguish between various kinds of collection, all being constituted of distinguishable but either more homogeneous or more heterogeneous entities, I alter the terminological conventions slightly. In what follows, I assume entities to consist of either non-discrete (e.g. *water*) or discrete parts (e.g. *team*). This distinction is not necessarily referential, but reflects human construal of these entities. That is to say that, although a substance like water, but also an aggregate like rice consist of (very) small particles (atoms of water and grains of rice), they are too small to be conceived of as discrete. In contrast, a team and also other collections like a forest or furniture consist of entities which are bigger, which may move, which have different functions etc. and are thus salient enough to be conceived of as discrete. The notions of homogeneity and heterogeneity will only be used to additionally categorise discrete entities as being similar to each other or not. With this in mind, the classification of water as being a non-discrete and homogeneous entity still holds: we cannot make out the single molecules of water, they are non-discrete and, as these are uniformly made of the same atoms, they are also homogeneous, i.e. similar to each other. The same is true for aggregates like rice, whose constituting grains are too small and too similar to each other to be con-

ceived of as discrete entities. However, especially in the case of collections, the terminological shift allows for a distinction to be made between collections made of either homogeneous (e.g. *a bouquet of very similar flowers*) or perceptually and/or functionally heterogeneous (e.g. *furniture, team*) entities. In both cases, the entities are distinguishable but differ in their degree of homogeneity, i.e. similarity. Factors influencing this degree of similarity may be animacy, functionality, as well as perceptual aspects. The constituting entities of furniture are thus conceived of as heterogeneous because they are perceptually different and they all have a different function. In a similar way, although the members of a team are perceptually homogeneous because they all wear similar outfits, they may rather be conceived of as heterogeneous, since they are animate and thus move individually and they all have different functions. The examples given already suggest that the features of discreteness and homogeneity are gradable: grains of rice are more discrete than atoms of water, but to a lesser extent than pieces of furniture.

Summarising, there are extra-linguistic entities and events (like items of clothing, leaves on a tree or playing the flute) which we can describe using different linguistic means of expression. These means reflect how we interpret what we perceive, which elements we focus on and which ones we neglect. These mechanisms are not exclusive to the verbal or the nominal domain, but are based on general cognitive principles such as attention, perspective or focus (cf. Croft/Cruise 2004; Langacker 2008). Jackendoff (1983, 42) sums this up:

One of the most obvious aspects of the projected world is that it is divided up into #things# – #entities# with a certain kind of spatial and temporal integrity. In the simplest case, a #thing# is the figure of a figure-ground opposition in the visual field; by contrast with the figure, the ground is unattended and relatively less vivid. In more complex cases (such as ordinary life), a multitude of #things# are perceived in the visual field, standing or moving in various relations to one another.

Terminologically, the diverse theories roughly differ in adopting the notion of *construal* (Langacker, Croft/Cruise, Jackendoff) or some other terms like *cognitive ambiguity* (Ising 2019), *apprehension* (Seiler 1986), *Seinsart* (Rijkhoff 2002) or *Nominalaspekt* ‘nominal aspect’ (Meisterfeld 1998).

On the surface, all these approaches to how we linguistically interpret the world seem to be very similar. In contrast, example (3) indicates a rather important difference between them: Jackendoff (1991a, 19–20) speaks of *entity construal* typically expressed by certain nouns (e.g. mass vs. count nouns), but uses both nouns and full noun phrases as exemplars. Contrary to that, Rijkhoff (2002) explicitly speaks of types of noun phrase that are characterised by having different features in different languages and Croft/Cruise (2004) and Langacker (2008) stress the conceptual nature of boundedness and discreteness, special noun types being a mani-

festation of their underlying cognitive basis. Meisterfeld (1998, 34) defines nominal aspect as a grammatical category and mainly focuses on its morpho-syntactic aspects. Finally, Ising (2019) adopts the most global perspective in considering a number of different grammatical and lexical means of expression of reconceptualising an extra-linguistic entity, be it certain nouns, anaphoric reference, the article system of a language and so on. A second important difference has already been mentioned. Whereas purely cognitive approaches concentrate mainly on the English language, Rijkhoff, Seiler, Meisterfeld and Ising focus on language comparison (on a larger or smaller scale). These issues can be subsumed under the question of whether the features of boundedness and discreteness are viewed as a fixed part of the internal conceptual and/or semantic structure of a noun, whether they are rather grammatical features expressed for instance by article use as in example (3), or whether they are features independent of the language that expresses them.

To address this issue, I embark on a detour to the verbal domain where the very same question is also raised. In a parallel way, nominal as well as verbal aspectuality are onomasiological categories that describe the possible linguistic means of expression of the internal structure of an entity or an action as well as its outer boundaries – both features describe how it is construed (cf. ex. (1) and (2)). The encompassing term *verbal aspectuality*, as used by Dessì Schmid (2014; 2019), is traditionally divided into uni- and bi-dimensional approaches. The latter differentiate between the two major domains of semantic means of expression – *lexical aspect* or *Aktionsart* – and grammatical ones – *grammatical* or *verbal aspect* (Sasse 2002; Boogaart 2004; Boogaart/Janssen 2010, 813; Filip 2011; Dessì Schmid 2014; 2019).

The concept of *Aktionsart* as it is widely adopted today mainly goes back to Slavonic tradition and Zeno Vendler, who used the term *time schemata* (cf. Vendler 1957); the notion of *Aktionsart* had already been introduced in the late 19th century as a term covering both lexical and grammatical aspect (cf. Boogaart 2004, 1167). Vendler elaborated a typology of verbs and entire predicates that describes how these linguistically structure a specific situation. Based on his seminal article, we generally differentiate four *Vendler classes* constituted by the three dichotomies *telic–atelic*, *dynamic–stative* and *durative–punctual*. Smith (1991) furthermore adds the category of semelfactives which are dynamic, punctual and atelic (e.g. *to cough* or *to kick*) (cf. Table 1.1).

Central to this concept of *Aktionsart* is that it is anchored to the verb. Consequently, verbs like *to stop* refer always to achievements or verbs like *to hate* are always states, no matter in what tense or (grammatical) aspect they are presented (cf. *infra*). This is why *Aktionsart* is often called *lexical aspect*, because it is assumed to be inherent to the verb. Note, however, that Vendler also treats VPs

Table 1.1: Aktionsarten according to Vendler (1957) (table adapt. from Boogaart 2004, 1169).

	dynamic	telic	durative	example
state	–	–	+	<i>to know</i>
activity	+	–	+	<i>to run</i>
accomplishment	+	+	+	<i>to recover</i>
achievement	+	+	–	<i>to find</i>

and not just verbs. For instance, he explains the difference between activities and accomplishments with the difference between *to run* or *to draw*, which are both atelic, and *to run a mile* or *to draw a circle*, which are telic and thus accomplishments (cf. Vendler 1957, 145–146; cf. also Boogaart 2004, 1166–1167).

In contrast to the traditionally more lexical concept of Aktionsart, verbal aspect is defined as “ways of viewing the internal temporal constituency of a situation” (Comrie 1976, 3) and is seen as the (grammatical) expression of an action being represented as completed or as still ongoing (cf. also Bertinetto 1986). This distinction between perfectivity and imperfectivity varies from one language to the other, and typologically, we find various morphological, syntactic or lexical means of expression of verbal aspect (see Comrie 1976, chap. 5 for an overview). As shown in example (4), Romance languages, for instance, express aspect in the past tense by means of inflection, whereas in English only imperfectivity can be overtly and unambiguously marked through the verbal periphrasis *was eating*; *ate* alone is not marked for aspect (examples from Dessì Schmid 2019, 10; cf. also ex. (2); see Bertinetto/Squartini 2016 for an up-to-date overview of the Romance tenses and their aspectual values):

- (4) Leo mangiò_{perf.}/mangiava_{imper.} un cornetto al cioccolato.
 ‘Leo ate_{perf./imperf.}/was eating_{imperf.} a chocolate croissant.’

This brief overview illustrates some issues also valid for the nominal domain. Firstly, Aktionsart is in general defined as a lexical category, but a verb as part of a syntagma can easily change its constitution, as we have seen above (cf. *to run* vs. *to run a mile*). Furthermore, different languages do not express Aktionsart and aspect necessarily with lexical and grammatical means respectively. So, an Italian speaker has to choose between the perfective or the imperfective inflection when s/he uses the past tense, as exemplified in (4). In contrast, a speaker of English or German has the option of choosing additional, more lexical means of expression of verbal aspectuality such as adverbials, as in *Leo aß gerade ein Croissant, als plötzlich die Katze auf den Tisch sprang* (‘Leo was just eating a crois-

sant when suddenly the cat jumped on the table'). Here, the adverbial *plötzlich* 'suddenly' marks the inflectional form *sprang* 'jumped' as perfective and the adverbial *gerade* 'just' marks *aß* 'ate' as imperfective; the inflectional forms are not marked for aspect in German. In this respect, lexical means of expression of aspectuality generally have a higher degree of optionality, while grammatical means show a higher degree of obligatoriness (cf. Lehmann 2015). Secondly, cognitive linguistics and recent grammaticalisation research indicate that there seems to be no clear demarcation between lexicon and grammar; rather, they are two poles of a continuum that underlies variation and change (cf. Comrie 1976, 6, footnote 1; Langacker 2006; 2008; Dessi Schmid 2019, 50–62). This overview indicated that a strict distinction between grammatical aspect and lexical Aktionsart can, if at all, only be maintained at the level of an individual language and "from a cognitive point of view, *aspect* and *aktionsart* [...] are actually one and the same thing, the difference being a matter of individual lexicalization and grammaticalization processes" (Sasse 1991, 32).

Adopting as a result of this a uni-dimensional model of aspectuality, I assume, following Dessi Schmid (2014; 2019), that aspectuality as a conceptual category describes how entities and situations are construed by language in space and time respectively. These means of linguistic expression can vary from lexical to grammatical, from one language to another and from one phase of language development to another. I therefore adopt a cognitive perspective on aspectuality as also argued by amongst others Croft/Cruse (2004) and Langacker (2008). This allows me to account for cross-linguistic variation, diachronic change and different means of expression of aspectuality.

Considering the numerous parallels between the verbal and the nominal domain, e.g. the incompatibility of unbounded mass nouns with bounded telic predicates – **to eat apple sauce in three hours* – (cf. i.a. Talmy 1988, 178–179; Krifka 1989; Brinton 1991; Jackendoff 1991a; 1991b, 27–32; Doetjes 1997, 44–56; Langacker 2008, 151–160) as well as the traditions of Romance philology (cf. Meisterfeld 1998), I adopt the notion of aspectuality also for the nominal domain. Coming back to the issues raised above, I thus assume the linguistic means of expression of nominal aspectuality to reflect the underlying construal of an entity. This structure may be either expressed by the semantics of a noun or the morpho-syntactic modifications the noun may be subjected to (cf. also the very similar approach of Ising 2019). In the verbal domain, these modifications are mostly represented by adverbs or inflection. In the nominal domain, aspectuality may equally be conveyed by lexical means of expression (various kinds of noun) or grammatical ones, e.g. by means of the indefinite article. Especially cross-linguistically, there are several modifiers which can change the aspectual constitution of a noun. An example for these is, amongst others, found in Tagalog,

where the ambifix *ka-. . -an* changes the aspectual structure of an individual count noun like *pulo* ‘island’ to a collection noun: *kapulauan* ‘archipelago’ (cf. Gil 1996, 63–64; cf. also Rijkhoff 2002, chap. 4.2 for a typological overview of various means of expression of nominal aspectuality). In Romance languages, which are the focus of this study, such modifications are rare. Given, however, the existence of such typological variation, nominal aspectuality has to be considered as a conceptual category, whose language-specific expression varies from more lexical to more grammatical means. The constituting features of nominal aspectuality are boundedness and internal structure, which I have mentioned above and which I will now address in more detail. Following research on cognitive linguistics, I assume the feature of boundedness to express the construal of a referent as an individual (cf. e.g. Langacker 1987b; Wisniewski/Lamb/Middleton 2003). As in the case of verbal aspectuality, the outer limits that define an entity as a bounded individual may be of various types. Consequently, in addition to prototypical *Spelke* objects (cf. Spelke 1994), a puddle, certain sounds, mental events or collections may all be linguistically construed as bounded entities (cf. Bloom/Kelemen 1995, 6–7; Langacker 2008, 141). An essential assumption here is that the feature of boundedness may not be totally present or absent but may, for example, be simply out of sight (cf. also Wisniewski/Lamb/Middleton 2003, 588): “[. . .], a speaker uses a –b[ounded] constituent to refer to an entity whose boundaries are not in view or not of concern; one can think of the boundaries as outside the current view. This does not entail that the entity is absolutely unbounded in space or time; it is just that we can’t see the boundaries from the present vantage point” (Jackendoff 1991a, 19).

This claim can be supported with coercion phenomena on the one hand and linguistic typology on the other hand. Type coercion in the nominal domain is typically associated with the *universal grinder* (count → mass; cf. Pelletier 1975), the *universal sorter* (mass → count; cf. Bunt 1985) and the *universal packager* (mass → count; cf. Bach 1986). All three phenomena describe either the focusing or weakening of boundedness. Whereas the principles may not be as universal as they are usually postulated to be – underlying idiosyncrasies and context restrictions (can a count abstract noun be grounded?) – the basic assumptions are perfectly justified. By using the universal grinder, we weaken the boundedness of an entity, this can be linguistically manifested by coercing e.g. a count noun into a mass noun. In (5a) the bench does not cease to be limited in space, but its boundaries are simply not of concern. In contrast, the feature of boundedness is focused via the universal sorter and packager, concentrating either on a type or a portion interpretation of the entity. Example (5b) refers to types of wine and in example (5c) the person wants a pre-defined portion of ice cream, like an ice lolly, and not an undefined mass:

- (5) a. You'll have to stand – there's not enough bench for another big person.
(Langacker 2008, 143)
- b. Hungary produces many excellent wines. (Bunt 1985, 10)
- c. I want an ice-cream. (adapt. from Bach 1986, 10)

These examples of type coercion illustrate that the feature of boundedness cannot be seen as a completely inflexible reflexion of grammar or semantics but must be assumed to be of conceptual nature which underlies changes of perspective displayed by linguistic means of expression, but also governed by pragmatics and world-knowledge. As shown e.g. by Lauwers/Vermote (2014, 164–165), speakers of French and Dutch are less likely to accept fruit nouns in mass syntax when presented in contexts other than cooking, where they thus naturally (still) appear in their discrete forms.

From a typological point of view, the mass-count distinction as a linguistic reflexion of boundedness is much debated. It seems to be a consensus that individual entities with time-stable properties are mostly labelled with count nouns and substances without outer boundaries with mass nouns – or equivalent grammatical structures (cf. e.g. Rothstein 2010a, 343–344). Apart from that, the question is far from settled as to why particularly supertypes¹ and granular aggregates (in the terminology of Grimm 2012) like *rice* or *gravel* are expressed by a count noun in language A and by a mass noun in language B. For instance, the concept of various precious objects a person wears for adornment is expressed in English, German, French and Italian with a heterogeneous mass noun (*jewellery*, *Schmuck*, *bijouterie*, *gioielleria*), but in Spanish and Portuguese with a count noun (*joya(s)*, *jóia(s)*) (cf. Wisniewski/Lamb/Middleton 2003 and Wisniewski 2010 for an overview of various studies analysing this question and possible influencing factors, such as the way we interact with the referents; cf. also Wierzbicka 1988). In both language groups, we refer to the same extra-linguistic entities, but Spanish and Portuguese focus on the boundedness of the single pieces of jewellery, whereas the other languages construe these objects linguistically as an unbounded mass – the discrete objects remain bounded entities, but they are simply not construed as such. Bale/Gillon (2020, 34) even go so far as to assume that the relation between mass and count nouns and their real-world equivalents is no more motivated than other features like animacy and gender.

¹ I use this notion, originally coming from database management, to refer to superordinate categories in general. In anglophone literature on this topic, superordinate categories are typically used to refer to categories verbalised by a count noun, we thus often find the distinction between *superordinate categories* like *vehicle* and *mass superordinate categories* like *furniture* (cf. e.g. Wisniewski/Imai/Casey 1996). The notion of supertype thus stands for both count and mass superordinates.

Summarising, the feature of boundedness defines whether an entity is construed as having “some limit to the set of constitutive entities” (Langacker 2008, 136). This limit does not have to be concrete as in the case of *a pig* or *a committee* (cf. ex. (3)), but can also be more abstract as in the case of *hour* or *beep*. We consequently construe entities as bounded when we conceptually contrast them to other entities, when they have a certain internal configuration and/or when we see a special function in them (cf. Langacker 2008, 136–139). In number marking languages, bounded entities are generally labelled by count nouns and unbounded entities by mass nouns (e.g. *dog* vs. *water*).

Whereas the feature of boundedness tends to be taken as dichotomous, the feature of discreteness is seen rather as a continuum of different degrees. As already indicated above, a substance like water is characterised by being completely non-discrete and homogeneous, since we cannot distinguish any individual constituents. Granular aggregates like rice are still construed as non-discrete entities but, albeit to a lesser degree than water, we can still make out some small constituting, discrete but homogeneous entities. Finally, and intuitively, collections like a bouquet are constituted of discrete individuals that are, however, more homogeneous than those of a family or a team, where the constituting persons differ in various features. The factors influencing this construal as either homogeneous or heterogeneous, like e.g. animacy or functionality (cf. *supra*) result in two basic aspects: first, an entity is conceived of as homogeneous if its main properties and qualities are the same for every single portion of it. This holds for water, rice and the bouquet but not for a family or a bicycle. Second, an entity is conceived of as homogeneous when adding or taking away portions of it does not change the fact of its still being exactly this entity. Adding or taking away water or rice does not change its being water or rice. This does not hold for heterogeneous discrete entities like a team, as only three defenders and a keeper do not form a (functional) football team (cf. Langacker 2008, 139–142). These features of homogeneity are traditionally described by the notion of *cumulative reference* (cf. Quine 1960, 90–95). Like Langacker (2008), Meisterfeld (1998, 44–52) considers the internal constitution of a noun not as dichotomous but rather as a continuum between the two poles of discreteness and continuity. The referents of a count collective noun, for instance, are more coherent than those of the inflectional plural, but the referents of a typical mass noun – Meisterfeld (1998, 52) calls them *Kontinua* ‘continuous nouns’ – show a higher degree of cohesion than those of collectives (cf. also Gil 1996, 63; Acquaviva 2008, 101–105). Rijkhoff (2002, 51–53) also assumes the feature of homogeneity to be crucial for a classification of noun types. In contrast to Langacker and Meisterfeld, he sees the feature as dichotomous, however. A very important aspect stressed by Rijkhoff is that homogeneity may not only become evident by means of reference, but also by linguistic means

of expression. The distinction between [\pm homogeneous] nouns may be e.g. manifested by different kinds of nominal classifiers, viz. sortal or mensural ones. In a language like Thai, homogeneous or non-homogeneous nouns are thus distinguished by the obligatory use of a mensural classifier for the former and a sortal classifier for the latter type. A language like Yucatec Maya, on the other hand, does not distinguish between these two types of classifier and thus encodes all nouns independently of the feature of homogeneity (cf. Rijkhoff 2002, 46–49; cf. *infra* for a detailed discussion of classifiers as linguistic means of expression of nominal aspectuality). As in the case of the aspectual feature of boundedness, the construal of extra-linguistic entities as being either non-discrete or discrete consequently depends on focusing on or neglecting certain aspects. Granular aggregates like gravel or sand are thus construed as internally homogeneous although they are constituted of more or less clearly distinguishable parts (cf. Chierchia 1998a; cf. also Middleton et al. 2004 for psycholinguistic evidence). It may thus be useful to distinguish between homogeneous and heterogeneous as well as discrete and non-discrete entities to categorise different kinds of internal configuration. A team and a bouquet are then both constituted of discrete referents, but they are heterogeneous in the case of the team and homogeneous, i.e. more similar, in the case of the bouquet.

In most cases, internal continuity and external unboundedness as well as internal discreteness and boundedness correlate – substances are mostly expressed by mass nouns and distinctive entities by count nouns. This does not always have to be the case, as count nouns like *lake* or object mass nouns like *furniture* show (cf. Langacker 2008, 141); the latter will be treated in more detail in chapter 3.1. Summing up, “although categories may be similar in that their extensional units involve multiple entities, the cognitive agent takes an active role in construing those multiple entities. As a result, the cognitive agent may construe multiple entities in different ways, leading to different types of categories” (Wisniewski/Clancy/Tillman 2005, 125).

Nominal aspectuality is thus the linguistic reflection of this cognitive agent in construing extra-linguistic entities as either bounded or unbounded and as being somewhere between non-discrete and discrete. How this reflection is specifically expressed via linguistic means of expression will be treated in the next chapter.

1.2 The linguistic expression of nominal aspectuality

The previous section elaborated the conceptual category of nominal aspectuality, the features that constatae it and the implications that follow from it for cross-linguistic comparison and diachronic change. In the following, I give an overview

of possible means of expression of nominal aspectuality, first from a typological point of view and then specifically for Romance languages.

1.2.1 The typological perspective

As already indicated above, the feature of boundedness is typically associated with the mass-count distinction, boundedness being a prerequisite of countability (cf. Meisterfeld 1998, 40). It follows from this that count nouns reflect boundedness and mass nouns unboundedness (cf. Jackendoff 1991a, 19; Meisterfeld 1998, 26; Langacker 2008, 131). Using count or mass nouns is, however, only one possible option, specific for number marking languages. In this respect, Chierchia (2010, 107–108) distinguishes three language types according to their coding of mass vs. count syntax. Firstly, there are the Indo-European languages, for example, which mostly mark mass and count via obligatory nominal number marking and determination. Secondly, languages like Mandarin Chinese do not mark nominal plurality morphologically, but code the mass-count distinction by means of different classifiers. Those are linguistic elements which categorise referents so that they can be counted, measured etc. and appear mostly in Asian languages, but also in Africa and the Americas (cf. Grinevald 2004, 1016; 1021). Thirdly, in languages like Tagalog or Dëne Sųłine (Na-Dené language family, spoken by about 12,000 people in north-western Canada), there is neither obligatory number marking nor an obligatory classifier system; numerals can be directly combined with a number neutral noun. They, however, still distinguish count vs. mass via the insertion of measure terms only possible with mass nouns (cf. also Doetjes 2011, 2564).² Before coming to a more detailed description of these systems, a note on the presence of nominal number marking in general is necessary: Greenberg (1963, 96) argued that there are no languages in the world that do not dispose of a numerical distinction – at least in the pronominal system. We now, however, know of some counter-examples against this generalisation: Pirahã, for instance, spoken in the Amazonas region in Brazil (220 speakers were counted in 1997), makes use of personal pronouns in the first, second and third person, but these can express both singular and plural. Old Javanese is supposed to have functioned in a similar way and Classical Chinese also seems to have been devoid of nominal number

² One may differentiate an additional fourth type. As shown by Oliveira de Lima (2014), the language of Yudja, spoken by about 300 people in the Xingu Indigenous Territory in Brazil, allows for a direct combination with numerals and bare nouns, both in the case of nouns denoting individuals and of substances. As this language however seems to be a special case, it will not be treated any further in what follows.

(cf. Corbett 2000, 50–51; Iturrioz-Leza/Skopeteas 2004, 1053). Other languages like Maranungku (Western-Daly, North-Australia) do not encode number in the noun, but mark it in the verb (cf. Dryer 2005, 139). All in all, Dryer (2005) enumerates 98 languages that do not encode nominal plurality, as opposed to 968 that do encode it in some way. Note that these former languages mostly share the same pattern as Maranungku. With respect to the latter, Corbett (2000) develops a typology of number marking in the languages of the world. These can be mainly of three types. Either a language has two distinct forms for the singular and the plural – this is the case for Indo-European languages – or they additionally use a *general number*, this being unspecified for singular or plural. In this case, there is the possibility for a language to have a special form for each of these number options. In Bayso (Afro-Asiatic, spoken in Ethiopia), for instance, the word for ‘lion’ is *lúban*, and this general number form can either refer to one singular individual or a plurality of lions – it is not marked for number. Additionally, *lubán-titi* ‘(a single) lion’ is explicitly marked for the singular and *luban-jool* ‘lions’ for the plural (cf. Corbett 2000, 10–11). In contrast to this, languages like Japanese have a shared form for the singular and the general number, but a special form for the plural. Accordingly, Japanese *inu* ‘dog’ can denote one single dog or more than one and *inu-tati* explicitly refers to a plurality of dogs, but there is no exclusive form for the singular as there is in Bayso (cf. Corbett 2000, 13–14). According to Corbett (2000, 16–17), there is the possibility of a fourth system with the plural and general number sharing a form and the singular being marked, but a natural language like this does not seem to exist. Languages may additionally differ in the values they ascribe to the category of plurality. Either they only oppose a category of ‘one’ to another of ‘more than one’, or they additionally distinguish other plural values like the dual, the trial or the paucal (cf. Corbett 2000, 20–26), though these categories are not greatly relevant for the present study.

Strictly speaking, the three language types referred to by Chierchia all belong to Corbett’s number marking languages. Even in classifier languages like Mandarin (type 2) or number neutral languages like Tagalog (type 3) number marking is not absent, but only optional (mostly with animate or human nouns, cf. *infra*), they share the pattern of Japanese. A Mandarin general number noun like in (6), for instance, can express both the singular and the plural, but it is the context that determines the actual number of the referent (cf. Corbett 2000, 13–16; Doetjes 2011):

- (6) Shi lǘ shū
 be green book
 ‘It’s a green book.’/‘They are green books.’ (Wiedenhof 2015, 249)

However, both languages, Mandarin Chinese and Tagalog, can overtly code the mass-count distinction (despite the fact that they mark number only optionally). Type 2 languages use an obligatory system of classifiers which have to be inserted when one wants to count or measure a noun. These classifiers are usually restricted to one or another noun type, as in Mandarin *wǎn* ‘bowlful’, *kǒu* ‘mouthful’, *bǎ* ‘thing wielded’ or *ge* ‘item’ (cf. Wiedenhof 2015, 274–278). These examples, together with some other restrictions like the presence of the modification marker *de* which is only possible with mass classifiers, suggest that languages like Mandarin have a mass-count distinction similar to that of English, for instance (cf. Cheng/Sybesma 1998; Li/Barner/Huang 2008). Finally, type 3 languages like Tagalog have neither an obligatory number marking nor an obligatory classifier system. The noun system behaves, however, basically like the English one, whereby typical count nouns like *mansanas* ‘apple’ can be directly combined with numerals, and typical mass nouns like *bigas* ‘rice’ have to be accompanied by a container word to be counted (cf. Chierchia 2010, 108; Doetjes 2011, 2564). It may thus be summarised that the mass-count distinction as a linguistic reflection of the feature of boundedness correlates with obligatory number marking in Chierchia’s type 1 languages, but may also be expressed by other linguistic means. Even the obligatory plural marking can be expressed very differently, whether through plural suffixes as in the case of Indo-European languages, reduplication (in most of the Australian languages), prefixation (as in Swahili) or prosodic means (as in Shilluk, Nilo-Saharan) (see Iturriz-Leza/Skopeteas 2004; Dryer 2005 for an excellent overview).

Another means of expression of boundedness is definiteness. This can, for instance, be marked in German by the definite article. In example (7a) the mass noun *Wasser* ‘water’ is a bare noun; without any determiner it is indeed interpreted as an undefined mass. In contrast, in example (7b) the definite article *das* binds this undefined mass and triggers the inference of a typical container for water taken to the kitchen, maybe a bucket (cf. Meisterfeld 1998, 43):

- (7) a. Er holte Ø Wasser an der Pumpe.
 ‘He got water from the pump.’
 b. Er brachte das Wasser in die Küche.
 ‘He took the (bucket of) water to the kitchen.’

At first sight, this postulation seems to be contradictory, given the classification of Chierchia (1998a, 56), who clearly categorises the definite article (in English) as being neutral to the mass-count distinction, and thus aspectual bounding. Indeed, the definite article often serves to simply create a relation between a noun and the extra-linguistic entity which it names. Coseriu (1955, 36–37) calls this *actualización* ‘actualisation’ which he defines as the transformation from a

potential to a real designation, i.e. denotation. This is most true for the English article, which does not indicate gender or number, unlike its Romance counterparts or the definite article in German. This simple actualisation process does not necessarily coincide with other determination processes like individualisation or quantification, but only indicates that something exists (cf. Coseriu 1955, 38). With mass nouns like in example (7), however, the definite article, at least in languages like English, German and Romance languages, serves as a contextual binder. As such, the definite article may be compared to the sortal classifiers in Mandarin Chinese. This is well illustrated by a number of other examples like *Could you pass me the salt* (= the bounded amount of salt available in the shaker) or *The furniture is modern* (= the bounded amount of furniture in the room or the house in question). An exception of this function of the definite article are generic contexts, which in Romance languages are also marked by the definite article, where it does not add contours: *Le sel est un minéral* ‘Salt is a mineral’ (cf. Meisterfeld 1998, 41). Similar characteristics may be determined for possessive and demonstrative pronouns, “[. . .] puisque leur caractéristique fondamentale est de présenter l’entité dénotée, quel que soit son type, comme une entité borne, qui a ses limites propres déjà établies, [. . .]” (‘since their fundamental characteristic is to present the denoted entity, whatever its type, as a bounded entity, which has its own limits already established’) (Kleiber 1998, 93).

Summarising, the grammatical mass-count distinction as a linguistic reflection of conceptual (un)boundedness can be expressed by very different means in languages around the world. Additionally, as we have seen, number marking does not always have to be obligatory. With respect to the obligatoriness of number marking, many languages differentiate between inanimate, animate and human referents. Hatam (Western Papua, Indonesian) for example mostly has only number neutral nouns, but nouns referring to humans can be optionally marked via the clitic *-nya* to express plurality; such a marking of number is obligatory in languages like Jamul Tiipay (Yuman, southern California) and some languages even differentiate between kinship terms and terms for other human beings (e.g. Amele, Trans-New Guinea, Papua New Guinea) (cf. Comrie 1981, 182–183; Haspelmath 2005, 142). Mandarin Chinese furthermore makes use of an optional plural suffix *-men* for human nouns which additionally groups the referents together and relates them to a central individual (cf. Wiedenhof 2015, 301–303). The level of animacy not only affects the obligatoriness of number marking, but influences a whole range of linguistic phenomena and is typically captured via a scale or hierarchy. Such an animacy hierarchy was first introduced by Silverstein (1976). He examines different case-marking systems and finds that in split-ergative languages more animate subjects follow the nominative-accusative schema and lesser animate subjects the ergative-absolutive schema. Languages differ in

where they actually draw the line between the two schemata, but the hierarchy is always the same. Definiteness additionally seems to influence case marking in these languages, personal pronouns > proper names > NPs etc. are consequently situated higher in the hierarchy than the respective following category. More recent works generally simplify Silverstein's detailed model and we mostly find versions of the animacy and definiteness hierarchy similar to these (here adapt. after Aissen 2003, 437):

- (8) a. Animacy scale
 Human > animate > inanimate
 b. Definiteness scale
 Personal pronoun > proper name > definite NP > indefinite specific NP >
 indefinite unspecific NP

Apart from case marking in split-ergative languages, these hierarchies also play a role in differential object marking (cf. Aissen 2003), (verbal) agreement phenomena (cf. Kuno/Kaburaki 1977; Lehmann/Moravcsik 2000, 734) and, as described above, nominal number marking. To what extent animacy can also influence the linguistic behaviour of collection nouns will be addressed in chap. 4.2.2. Typologically, the presence or absence of number marking may also depend on reference, case or gender (cf. Rijkhoff 2002, 106–119 for an overview). Summarising, although many languages do not use an obligatory number marking system, they do encode nominal aspectuality in some way or another. This is mostly reflected by the presence or absence of boundedness and consequently countability.

Typologically, the feature of discreteness or internal structure seems to be less relevant with respect to the overt coding of nominal aspectuality, so there is thus no uniform grammaticalised principle similar to plural marking in the domain of boundedness. Rijkhoff (2002, 54–55) assumes in this respect that spatial orientation, via e.g. boundedness, is more primary to human cognition and consequently more prominent in language than the linguistic expression of an internal structure. Whether or not a noun refers to discrete individuals or rather to a non-discrete substance as part of extra-linguistic reality may thus be expressed rather by more lexical, non-obligatory means of expression. For instance, one may speak of *furniture* and *water*, embedding the two respective nouns in mass syntax and thus expressing their unboundedness, without speaking of their internal structure. The same applies to nouns denoting bounded entities like *boy* or *team*, where we are usually uninterested in the constituting parts of the denotatum we are referring to. The internal structure of an entity may nevertheless be linguistically accessed. One means of expression is, for instance, a category of predicates that Schwarzschild (2011) calls *stubbornly distributive*. Adjectives like *round*, *small* or

slim are only compatible with nouns denoting entities that have an internal structure. Consequently, *round furniture* is perfectly fine, but **round water* is ungrammatical since the noun does not denote discrete entities which may be predicated with *round*. With their stubborn distributiveness, these kinds of adjective are thus sensible to the feature of discreteness. They are, however, also only optional. We do not have to specify whether the furniture we bring upstairs is round or square, only if we have to additionally concretise its properties. One may imagine, for instance, a sentence like *The round furniture does not fit through the door, we have to move it through the big window*. My impression is furthermore that the feature of discreteness is less relevant for cross-linguistic variation, and thus more dependent on the actual denotata and not their construal – that is, it does not matter what kind of word any language in the world uses to verbalise tables and chairs, they may always be predicated as being round. There has, however, to my knowledge been no systematic study on this. With this example, a specific connection between boundedness and discreteness may already be indicated: the internal structure of an entity denoted by some noun may only be linguistically accessible when this entity is not bounded, otherwise its outer frontiers prevent the access. *People* may thus be tall, but *tall team* is weird, since the boundedness of *team* hinders the linguistic access to the constituting entities (cf. Borillo 1997, 113; Flaux 1998, 179–180; Bosque 1999, 54; Mihatsch 2000, 56–58). This aspect will be discussed in more detail in chaps. 3.1.2 and 4.2.2. Consequently, I will not elaborate a typological overview of the linguistic reflection of discreteness, but will instead come to this point when I specifically address this feature with respect to Romance languages. This is also the reason why I will concentrate purely on the feature of boundedness in the following chapter.

1.2.2 Nominal aspectuality in Romance languages

As illustrated in the last section, nominal number marking in the languages of the world is far from being a straightforward phenomenon – how do Romance languages fit into the picture? Like English, Romance languages have an obligatory number system with different forms for singular and plural – they do not make use of a general number category. Consequently, very roughly speaking, the singular expresses ‘only one’ and the plural ‘more than one’,³ and count nouns can

³ Evidently, this is a strong simplification of number marking in Romance languages. For instance, generic contexts do not fall into this classification, because Romance languages like English or German can also use the singular to refer to a whole class, i.e. a plurality (cf. Kleiber 1990, 19; Corbett 2000, 19). In addition, the inflectional plural may also express the absence of

be directly combined with numerals. Other ways of expressing nominal aspectuality in the Romance languages will be discussed in what follows.

Although the generalisation of the obligatoriness of plural suffixes may be true for all Romance languages, there are rather striking differences between them. A first approach lies in the traditional distinction between Western and Eastern languages which differ either in forming the plural with a final *-s* or through vocalic endings as residues of either the Latin accusative or nominative plural (cf. Wartburg 1936). Following Maiden (2016) I do not regard French as belonging to the Western languages, as it has traditionally been categorised. I rather assume that it forms a category apart, because plural and also gender are mostly not expressed phonetically in the noun, but only in prenominal determiners and modifiers, as the following example shows:

- (9) un ami/une amie – des ami(e)s
 [œnami/ynami – dezami]
 ‘a male friend/a female friend’ – ‘male or female friends’ (Stark 2008, 51)

In the light of this, I therefore assume with respect to nominal number marking some kind of invariant coding in French, unambiguous plural marking with noun final *-s* in the Ibero-Romance languages⁴ and relatively ambiguous vocalic suffixes in Italian and Romanian (cf. Maiden 2016, 697–698; Pomino 2017, 704–707). The ambiguity, particularly in Italian, lies in the form syncretism of some suffixes like *-e* which can express both feminine and masculine singular (e.g. *studente* masc. sg. ‘student’) as well as the feminine plural (e.g. *case* fem. pl. ‘houses’) or *-a* which expresses both feminine singular (*casa* fem. sg. ‘house’) as well as the collective plural in some nouns designating mostly body parts (*braccio* masc. sg. ‘arm’ – *bracci* masc. pl. ‘arms (fig.)’ – *braccia* fem. pl. ‘arms’) (cf. Schön 1971, 83–121; Stark 2008, 51–53). Summarising, the Romance languages express boundedness through the more or less overt and unambiguous morphological marking of plurality. Unboundedness however is mostly the (grammatically) unmarked option and only manifests itself through the absence of possible plural forms or articles (apart from sortal or packaging readings, cf. *supra*; cf. Stark 2009, 285).

plurality, especially in the case of existential sentences like Fr. *il n’y avaient pas d’enfants*_{pl.} *dans la rue* (‘there weren’t any children in the street’).

⁴ In some Spanish varieties, deletion or aspiration of final *-s* is compensated through opening or lengthening of word-final vowels. They too represent a problematic case for the traditional East-West split of the Romania (cf. Pomino 2017, 705).

There is, however, another means of expression of nominal aspectuality in Romance languages which does not lie in number but in the gender system, more precisely in the apprehending potential of the neuter gender.⁵ There is on the one hand the already mentioned possibility of a collective plural found, for instance in Italian. The collective plural *-a* is etymologically not an equivalent of the feminine gender, but stems from the Latin neuter plural. In Italian, it expresses the aspectual feature of boundedness in adding some delimitation to the plural: Italian *le braccia* does not mean ‘the arms’, but ‘both arms of a person taken together’ (cf. Schön 1971, 84–94). On the other hand, some Romance varieties (Asturian, Cantabrian, some central-southern varieties of Italo-Romance) make use of the so-called *neutro de materia* ‘substance neuter’ which morphologically marks the respective noun as unindividuated (see Loporcaro 2016, 933–934; Pomino 2017, 695–697 for an overview). This not only holds for typical mass nouns like Borghese (Lazio) *ferro*_{mass} ‘iron’ vs. *ferro*_{count} ‘iron implement’, but also for abstract nouns and nominalised adjectives (cf. Kučerová/Moro 2011). Some varieties not only code this distinction in noun morphology, but additionally in the determiners and modifiers. Central Asturian, for example, overtly marks the mass neuter not only in the noun itself, but also in post-nominal adjectives (cf. (10); cf. Hualde 1992; Fernández-Ordóñez 2009; Tuten/Pato/Schwarzwald 2016, 398; Loporcaro 2018, 160–194), whereas Central Italian only distinguishes between a mass definite article *lo* and the masculine and feminine count variant *lu* and *la* (cf. Loporcaro 2018, 132).

- (10) el pilu_{count} blancu_{count} – el pelo_{mass} blanco_{mass}
 ‘the (single) white hair’ – ‘the white hair’ (Pomino 2017, 696)

An additional way of expressing (un)boundedness in Romance languages is the determiner system. As in the case of the classifier system in Mandarin, Romance languages also have a nominal categorisation system which marks the aspectual feature of (un)boundedness. Firstly, all Romance languages have an indefinite article that stems from the Latin numeral *ŪNUS* ‘one’ and marks nouns as bounded (cf. Stark 2008, 47; Vincent 2017, 740–741). Whereas every Romance language has a prenominal determiner marking boundedness, only some of them also make use of a determiner marking unboundedness. Apart from the already mentioned possibility of mass neuter agreement in Asturian, Cantabrian and Central Italian, only French and Standard Italian are characterised by the exist-

⁵ For the apprehending potential of the Latin gender system see Schön 1971; Stark 2007, 55–56; 2008, 49–50; chap. 3.2.

ence of a so-called partitive article. The presence of this kind of mass-denoting determiner correlates, according to Stark (2008), with the absence of an unambiguous plural marking system in the noun. Recall that Italian and French are those Romance languages lacking such a system; in contrast, the central-southern Italian dialects with a very differentiated gender system (cf. *supra*) do not use a partitive article (Stark 2008, 55, footnote 9). The use of the partitive is illustrated in the following example (adapt. from Stark 2008, 46):

- (11) a. Compro pão. (Portuguese)
 b. Compro pan. (Spanish)
 c. Compro (del) pane. (Italian)
 d. J'achète *(du) pain. (French)
 'I buy (some) bread.'

In Spanish and Portuguese mass nouns appear without any determiner in argument position (cf. *infra* for the possibility of bare nouns in Romance languages), in Italian the partitive article *del* is optional and in French the partitive article is inserted before the noun.⁶ In the cases of French and Italian, the partitive article stems from the construction <*de*_{prep} + def. article> and in the medieval stages of these languages it served to designate a portion of some aforementioned referent. In this respect, it is restricted to concrete substances and to agreement position like in *del vin volentiers bevaient* 'they drank gladly of the wine' (*Erec et Enide*, 3178; cit. in Carlier/Lamiroy 2014, 493). During its grammaticalisation process it slowly extended to referents not previously mentioned in the context and abstract nouns to finally become a fully grammaticalised and obligatory indefinite determiner in French (cf. Carlier 2007; Carlier/Lamiroy 2014, 491–501). In Italian, the partitive article *del* undergoes the same changes; it also can be combined with abstract nouns in Modern Italian, but, in contrast to French, it is not obligatory – at least in northern Italian dialects (cf. *supra*; cf. Carlier/Lamiroy 2014, 506–514). In this respect, the two sentences in (12) are both absolutely fine, but not exactly synonymous: (12a) is said to still have some of its original, literally partitive meaning, whereas (12b) expresses an indefinite quantity of the wine (Carlier/Lamiroy 2014, 512).

⁶ As noted by Carlier/Lamiroy (2018, 151) the partitive article is still often missing with abstract nouns in prepositional phrases as in *Il prépare ce repas avec ∅ amour* 'he prepares this meal with love'. With a concrete noun, the partitive article has to be inserted, as zero-marking is exceptional with this noun type: *Il prépare ce repas avec du/?∅ chocolat/sel/sucre/lait* 'he prepares this meal with PART. ART./?∅ chocolate/salt/sugar/milk'.

- (12) a. Abbiamo bevuto del vino caldo.
 ‘We drank (some of the) mulled wine.’
 b. Abbiamo bevuto Ø vino caldo.
 ‘We drank (an indefinite quantity of) mulled wine.’

Summarising so far, the number system and its (non-)ambiguity as well as the determiner system define the nominal aspectuality feature of (un)boundedness in Romance languages. Finally, it is also the lack of determination which makes it possible to mark aspectuality. In Spanish, Italian and Portuguese bare singulars in argument position are only possible with mass and abstract nouns, whereby the lack of any article marks unboundedness (cf. (13a–c)). This is not possible in Modern French, where bare nominals are impossible in any syntactic position (cf. (13d); cf. also (11); cf. Stark 2008, 45–46; cf. also Kabatek 2008, 751 who speaks of a *negative grammaticalisation* of the zero article in the Romance languages):

- (13) a. Viste *(uma) águia_{count}? – Preciso de ø água_{mass}. (Portuguese)
 b. ¿Has visto *(un) águila_{count}? – Me falta ø agua_{mass}. (Spanish)
 c. Hai visto *(un) aquila_{count}? – Mi occorre (dell’)acqua_{mass}. (Italian)
 d. As-tu vu *(un) aigle_{count}? – Il me faut *(de l’)eau_{mass}. (French)
 ‘Did you see an eagle?’ – ‘I need (some) water.’

In subject position, bare singulars are not available in French, Spanish, Italian or European Portuguese, though in the latter three languages bare plurals are possible in post-verbal position (cf. Müller/Oliveira 2004, 16; Stark 2008, 46–47). In object position, there is in European Portuguese the possibility of singular nouns to refer to a plurality of entities, a phenomenon Meisterfeld (1998) terms *aspectual singular*. In these cases, a count noun is used in the singular with a non-distributive quantifier like *muito* ‘much’ or *tanto* ‘so much’ to stress its mass-like character. Examples like *muito carro parado* ‘(lit.) much parked car’ may thus be translated into English with something like ‘the mass of parked cars’ (cf. Meisterfeld 1998, 4). This syntactic operation that blurs the distinction between mass and count was also easily available in Old Spanish, but in the Modern varieties it has ceased to be very common (cf. Kabatek 2008). However, as noted by both Meisterfeld (1998, 9) and Kabatek (2008, 754–755), these constructions are only possible with a quantifier, and the use of the bare singular with plural reference does not seem to exist in European Portuguese and in (Old) Spanish. I have not come across any instances of aspectual singulars in Italian. Brazilian Portuguese does allow bare plurals as well as bare singulars in subject and object position, irrespective of the noun type or verb position. This is illustrated in (14a) and (14b),

where the bare singulars get a plural interpretation, boys and shoes respectively, as well as (14c), where *minhoca* ‘earthworm’ may either get a plural or a substance interpretation. The bare singular in Brazilian Portuguese thus seems to have a number neutral interpretation (examples taken from Müller/Oliveira 2004, 21–22; Pires de Oliveira/Rothstein 2011a, 2173; cf. also Schmitt/Munn 2002).⁷

- (14) a. Menino não pode entrar aqui.
 ‘Boys aren’t allowed in here.’
 b. Eu comprei sapato.
 ‘I bought shoes.’
 c. Essa lata tem mais minhoca do que aquela.
 ‘This can contains a bigger quantity of earthworm than that one.’

The overall picture of all these means of expression of (un)boundedness in Romance languages leads to the assumption of a continuum from French to Brazilian Portuguese. In French, the loss of overt and classificatory marking of gender and number has led to a system of obligatory determiners clearly marking a noun as either mass or count. In contrast, Brazilian Portuguese allows bare nouns in all syntactic positions and with both count and mass nouns; here the only possible means of differentiation between mass and count nouns is the incompatibility of mass nouns with the indefinite article or with numerals. Between these two poles lie European Portuguese, Spanish and Italian. European Portuguese and Spanish are very similar in not permitting any kind of bare singular in subject position and in only allowing bare plurals in post-verbal position, as in *Salen estudantes del edificio* ‘students leave the building’ (Stark 2008, 46). In these languages, bare singulars in object position however serve as markers for the mass-count distinction because they are only possible with mass nouns (cf. (11); Table 1.2). Italian behaves the same in this respect, but additionally has the not fully grammaticalised and therefore still optional partitive article to mark mass nouns as such. This overview is to be understood as a mere description of the language-specific characteristics and does not make any claim to be an explanatory elaboration. But, possible reasons for this distribution could lie in the ambiguity of nominal

7 There is an ongoing discussion about the actual character of bare singulars in Brazilian Portuguese, whether they are like mass nouns or whether they are disguised bare plurals. Furthermore, it is not exactly clear which kinds of referential function the bare singular may have, and whether kind-readings are possible or not (cf. Pires de Oliveira/Rothstein 2011, 2158 vs. Müller/Oliveira 2004, 23). This discussion is beyond the scope of this book and will not be further addressed (see Wall 2017; Gonçalves Rospantini 2018 for a summary and possible explanations).

suffixes marking number and gender as described above.⁸ The Romance continuum of the marking of the mass-count distinction is summarised in Table 1.2.

Table 1.2: The coding of the mass-count distinction in various Romance languages.

	French	Italian	Spanish	European Portuguese	Brazilian Portuguese
count DET	YES <i>un</i>	YES <i>un(o)</i>	YES <i>un</i>	YES <i>um</i>	YES <i>um</i>
mass DET	YES, abstract + mass nouns, obligatory <i>du</i>	YES, only mass nouns, optional <i>del</i>	NO	NO	NO
bare SG	NO	YES, only in argument position, only with abstract/mass nouns; no asp. singulars	YES, only in argument position, only with abstract/mass nouns; asp. singulars not common	YES, only in argument position, only with abstract/mass nouns; asp. singulars common	YES
bare PL	NO	YES, mostly postverbal	YES, mostly postverbal	YES, mostly postverbal	YES

⁸ There could be another explanation as postulated by Mulder/Lamiroy (2012), who determine a similar continuum of Romance languages (French > Italian > Spanish) with respect to the degree of grammaticalisation of aspectual periphrases, the subjunctive mood and demonstratives. Their explanation lies in the socio-historical circumstances of language change which is accelerated by urbanisation and mobility. In this respect, these processes happened fast in France from 1400 on, but rather slowly in Spain, with Italy occupying an intermediate position.

2 Interim summary: From nominal aspectuality to collection nouns

The last chapter elaborated the possibilities of construal of extra-linguistic entities and the linguistic means of expression of this construal. In many cases, the demarcation line between conceptual features and their linguistic instantiation is not easy to maintain, given the fact that there is often a one-to-one mapping of discrete objects labelled by count nouns and undefined substances labelled by mass nouns. Here, there seems to be thus an equivalent relation between the referent itself, its conceptualisation and its linguistic expression through a specific noun or noun phrase. When it comes to pluralities, however, things become a little more complicated. Before coming to the state of the art and the definitional delimitation of collection nouns, a more detailed delimitation of the possible kinds of plurality to be treated in the following remarks should first be made. In the introduction, I already introduced collection nouns as referring in their morphologically unmarked form to coherent pluralities of extra-linguistic entities. A plurality, however, can theoretically be understood as the flowers of a bouquet, pieces of clothing of an outfit, the members of a family, but also as grains of rice or sand or even atoms of water or particles of flour. As described in chap. 1, pluralities are present in all these examples (cf. also Chierchia 1998a), but they may just be out of conceptual and – supposedly representing it – linguistic sight. In terms of nominal aspectuality, all these nouns may thus be lined up on a continuum ranging from clearly discernible and heterogeneous people and pieces of clothing, more homogeneous flowers and lesser distinguishable grains of rice, to even more continuous grains of sand to the fully continuous substance of water. Given this continuum – and one can easily refine it by adding more steps – and my first definition of a collection, one may postulate that all these pluralities are cases of collections. Indeed, some authors also include granular aggregates in their definitional scope of heterogeneous mass nouns, and thus include examples like *oats*, *rice*, *wheat* or *gravel* since these, like *clothing* or *furniture*, refer to a plurality of entities verbalised by a mass noun (cf. e.g. Mihatsch 2016, 290); or as Wierzbicka (1988, 517) puts it: “Thus, rice can be seen either as an undifferentiated mass or as a collection of *grains* [. . .].” In the following I will, however, draw a demarcation line between those nouns which refer to clearly discernible objects, persons or animals and those whose plurality is constituted by granules or a substance. The reason for this clear definitional separation lies not only in the fact that “individual particles of rice or sand are of no special interest to people” (Wierzbicka 1988, 519), but also in the internal configuration of the plurality. Collections are to be understood as having some hierarchical order between the constituting entities

and the collection itself. This applies in the previously mentioned examples of Sp. *ropa* ‘clothing’ and Sp. *atuendo* ‘outfit’, where we find a hyponymic and a meronymic relation respectively: the trousers, the shirt and the shoes are either kinds of clothing or part of an outfit (cf. Cruse 1986). This is, however, not the case with granular aggregates like rice or sand: grains of rice are neither a kind of rice, nor are they a part of rice. Consequently, substances like *water* or *flour* are excluded from the following considerations as well as granular aggregates like *rice* or *sand*; these need only serve as reference points for comparison and delimitation. As we will see later, typical semantic domains of collection nouns, irrespective of their linguistic characteristics like countability, are thus particularly artefacts and persons, but animate collections like *herd*, *cattle* or *shoal* can also be found within this category. Attempting an interim summary that weaves together the two domains of nominal aspectuality and collections, I thus assume that nominal aspectuality is the linguistic expression, by lexical or grammatical means, of the construal of extra-linguistic entities. It is mainly defined by the two features of external boundedness and an internal configuration of an entity as non-discrete and homogeneous or discrete and either homogeneous or heterogeneous. Against the backdrop of this preliminary theoretical remark, collections are coherent pluralities of discrete entities that may vary in their outer boundedness (reflected by (un)countability of the verbalising noun) and their internal configuration (a collection may be made up of rather homogeneous or rather heterogeneous entities; either a hyponymic or meronymic relation between the collection and its constituting entities). Collections differ from granular aggregates in being organised in a hierarchical manner. In what follows, this definition of collections will be laid on the foundation of the state of the art.

3 Collection nouns: State of the art and definitional delimitation

The preceding two sections laid the general groundwork for a definitional delimitation of the category of collection nouns as it is understood in the framework of this present work. In the following, I will present the state of the art on different types of collection noun, bringing them together against the backdrop of the theory of nominal aspectuality described in chap. 1. In this respect, semantic-syntactic characteristics as well as morphological form are key. As already indicated in the introduction, collection nouns as a morphological category have up to now mostly been neglected, while research on their meaning and countability preferences has been more the focus of enquiry. The main assumptions here are that different types of noun that express a semantic plurality in their morphological base form can be subsumed under one umbrella term *collection noun* and that they only vary in gradual degrees. This basic theory will be supported by the following considerations.

3.1 Collection nouns as a semantic-syntactic category

As described in chap. 1, the linguistic construal of extra-linguistic entities can be well described with the theory of nominal aspectuality which comprises two main features, boundedness and discreteness. In this respect, typical bounded entities are expressed via count nouns and typical unbounded entities via mass nouns. Generally speaking, there is a correlation with boundedness and discreteness and unboundedness and non-discreteness. A bicycle, for instance, is normally labelled with a count noun because it is conceived of as a discrete entity composed of different parts which only altogether make up the object. A substance like water, however, is expressed with a mass noun, it is continuous and we can easily add or take away some of it and it still remains the substance under question – it thus refers cumulatively (cf. Quine 1960, 90–95). The correlations described above are straddled by the noun category focused on in the framework of this present work. First, count collective nouns (CCNs) like Fr. *bouquet* ‘bunch (of flowers)’ are generally countable, but are composed of a homogeneous set of discrete entities which implies the same properties of cumulativeness as with substances like water. Second, CCNs like Fr. *équipe* ‘team’ also consist of discrete entities, but in this case, they are more heterogeneous and we cannot just take parts away without changing the nature of the collection (cf. *supra*). In this respect, they would behave like individual objects such as the bicycle, with the difference

that the degree of cohesion of the constituting parts is higher in the latter case than in the former. Third, object mass nouns (OMNs) are not countable, but refer to a heterogeneous collection of discrete entities. Furthermore, typical CCNs are characterised by a meronymic relation between the collection and its parts, but the categorisation with respect to the semantic relations of OMNs and their constituting entities seems to be more difficult: is a chair a part of furniture or a type of furniture?

In what follows, I will present the state of the art on CCNs as well as singular and plural object mass nouns (SOMNs/POMNs) with respect to nominal aspectuality, focusing firstly on the criterion of external boundedness and secondly on the referential internal plurality of collection nouns. This approach aims to show that all these types of collection noun can be analysed under the same umbrella term, given the framework of nominal aspectuality. Before coming to this elaboration, a short note on the state of the art in general is necessary. There is an imbalance when it comes to research on CCNs and OMNs, with the former category being much more investigated than the latter. For CCNs, there is at present a huge amount of research which treats this category from various perspectives, linguistic traditions and in terms of the category's adoption by and adaptation to diverse languages so that an integral way of looking at CCNs seems to be impossible today, especially since there is no unanimously accepted definition of the notion of *collective noun* (cf. Gil 1996, 66–70; Meisterfeld 1998, 50). For instance, some authors also count those nouns as CCNs which only become such in discourse, e.g. through a metonymic shift⁹ (e.g. locatives like Fr. *le parterre* 'stalls' → 'people sitting there') (cf. i.a. Allan 1976; Borillo 1997; Flaux 1999; Lammert/Lecolle 2014, 218). Since the present work focuses on lexicalised collection nouns and therefore on aspects inherent to the respective noun, I will neglect this kind of phenomenon. This also holds for the Italian double plural (cf. e.g. *ginocchi* 'knees' vs. *ginocchia* '(both) knees'; *diti* 'fingers' vs. *dita* 'fingers (of the hand)') which differentiates a distributive and a collective plural in some nouns (cf. *supra*; cf. i.a. Schön 1971, 91–94; Ojeda 1995) and other 'false collective nouns' like Fr. *ville* 'city', which may denote collections of buildings and so on, but for which other semantic features are of higher importance (cf. Fasciolo 2016). Therefore, this look at the state of the art will not concentrate on Romance languages alone,

⁹ These shifts based on the principle of metonymy are very typical for these kinds of ad hoc semantic transfer, but also as mechanism of semantic change as permanent effects of these shifts. The underlying process is a figure-ground-change within the respective frame: the meaning of 'people sitting on the ground level of the theatre' switches from the ground to the figure, at the same time the original locative meaning switches from the figure to the ground (cf. i.a. Koch 1999).

but will also address research on English and German. With respect to the former, I will mainly focus on French because research on CCNs has not been the focus of interest for Spanish, Italian or Portuguese. I will furthermore try to give a complete picture in considering traditional grammars like that of Jespersen (1949) and modern cognitive semantic research undertaken, for example, by Lammert (2010). I will consequently present the state of the art of CCNs in French, completing it with work on English, German and the other Romance languages and looking at them from different theoretical angles.

In contrast, the state of the art on OMNs is much more uni-dimensional. First of all, there is no common terminology used in the literature. Apart from the notion adopted in this present work from Barner/Snedeker (2005), other terms are in use like *fake mass nouns*, *count mass nouns*, *individual mass nouns*, *collective mass nouns* or Fr. *noms de mass à référence hétérogène* ‘mass nouns with heterogeneous reference’ (cf. i.a. Wiederspiel 1992; Chierchia 2010; Doetjes 2011). Furthermore, Joosten (2006; 2010) advocates adopting the notion *aggregate* for mass nouns that refer to discrete entities: “An aggregate named *lingerie* or *bétail* is, then, more or less the same as the sum of its parts. That is also my motivation for choosing the term *aggregate*: etymologically *aggregate* (< Lat. AGGREGĀRE) means ‘sum’ result of an addition” (Joosten 2010, 42).

This definitional criterion of cumulative reference may hold for rather homogeneous discrete entities as in the case of *cattle* or *fruit*, whereas other categories named by nouns like *clothing* or *furniture* however seem to be in need of a degree of heterogeneity to be fully functional in certain contexts (cf. also Gardelle 2017). Put differently: if you want to furnish a room, it is not enough only to put tables and cupboards in it (cf. Grimm/Levin 2011). Consequently, within the category of OMNs, there seems to be a continuum of possible additive potential, with examples like *underwear* or *furniture* which seem to have lesser potential and *fruit* with more potential to refer cumulatively. As a consequence, the notion *aggregate* could be possibly adopted to mass supertypes like *fruit*, but not to all exemplars of this category. In the present work, I will adopt this notion only to granular aggregates like *rice*, *gravel* or *hair* which indeed are no more nor less than the sum of their parts (cf. also the argumentation of Wierzbicka 1988, 516–520 and my own considerations in chap. 2).

Whereas CCNs have been described and analysed since the beginning of the 20th century, research on OMNs is still in its infancy. On the one hand, English is almost the only language that has been investigated in this respect and, on the other, the majority of research has come exclusively from formal semantics (cf. i.a. Chierchia 1998a; 1998b; 2010; Rothstein 2010a; 2016b; 2017; Landman 2011). For Romance languages, OMNs usually merit only brief comments in the context of research on or descriptions of CCNs (cf. Bosque 1999, 53–56; Flaux 1999; Mihatsch

2006; 2016; Lammert 2010) or are merely treated as possible exemplars in related research areas (cf. i.a. Pires de Oliveira/Rothstein 2011a; 2011b who mention OMNs in the framework of bare singulars in Brazilian Portuguese).

3.1.1 External boundedness of count collective nouns and object mass nouns

As set out in chap. 1, an entity is construed as bounded when it can be perceived holistically, and this is linguistically reflected by countability. I will now examine in more detail the count or mass noun properties of different kinds of collection noun and their further implications, which will lead into a discussion of possible reasons for these syntactic features.

As already indicated, typical CCNs denote conceptually bounded entities (cf. Rijkhoff 1991, 296; cf. also Barker 1992, 69; Flaux 1998, 175; 1999, 473; Lecolle 1998, 46–47; Joosten 2006, 82–84; 2010, 43–44; Lammert 2010, 83–88; Lammert/Lecolle 2014, 209–210). As described in chap. 1.2, this feature is crucial for the application of the morphological plural, numerals and various determiners (cf. Meisterfeld 1998, 40). In Romance languages, this is represented mostly by a compatibility with the indefinite article which expresses countability and an incompatibility with the partitive article in French and Italian since the latter expresses uncountability (cf. Michaux 1992, 111; Flaux 1996–1997, 37; 1999, 479; Borillo 1997, 107; Mihatsch 2000; cf. also Table 1.2).¹⁰ The plural of a CCN thus does not express a plurality of entities, but a plurality of sets of entities. In the case of rather homogeneous entities, these outer boundaries of a CCN may be erased and a collection can be separated in two (*la division d'un troupeau* 'the separation of a herd') or two collections can be fused into one (*on fera fusionner les troupeaux en un seul* 'we will fuse the herds into a single one') (Flaux 1998, 174; 1999, 474–475; 484; cf. also Borillo 1997, 107–108; Lecolle 1998, 58). Another linguistic consequence of the boundedness of CCNs is the incompatibility with certain adjectives which qualify individual, mostly physical properties, like in Fr. **le jury est gros* 'the jury is fat' (cf. Borillo 1997, 113; Flaux 1998, 179–180). Another option is that these kinds of adjective only modify the whole collection, but not the constituting entities, like

¹⁰ Flaux (1999, 481) assumes that some collective nouns like Fr. *famille* 'family' can also be combined with the partitive article although they otherwise share all characteristics of a CCN. She deduces from this that "la clôture des N col. n'entraîne donc pas automatiquement le trait + comptable" ('the boundedness of collective nouns does not necessarily lead to the feature + countable'). Nevertheless, the contexts in which these kinds of structure are felicitous are limited and rather specific like *avoir de la famille en Normandie* 'to have family/relatives in Normandy' (cf. Flaux 1999, 482–483).

in Sp. *gentío grande* ‘big crowd’ (cf. Bosque 1999, 54; Mihatsch 2000, 56–58). In both cases, the individuals that make up the collection are not accessible, as the ‘outer frontier’ of the CCN hinders their modifying by these kinds of adjectives. Joosten (2010, 43–44) calls this focus on the outer boundaries *set profiling*. The principle of contiguity binds the collection by spatio-temporal, functional or institutional proximity (cf. Joosten 2010, 34; Lammert 2010, 87).

SOMNs on the contrary are *per definitionem* not countable and not combinable with the indefinite article; syntactically they behave like homogeneous mass nouns like *water* (cf. i.a. Chierchia 1998a, 55–56; Gillon 1999; Rothstein 2010a, 346).¹¹ As a consequence, they cannot be put into the plural and we need a classifier to linguistically access one single piece of the collection:

(15) *three furnitures/three pieces of furniture (Rothstein 2010a, 376)

The question that has puzzled researchers the most so far with regard to this aspect is the exact reason why we cannot count SOMNs.¹² Since they imply various discrete entities, we should be able to count these, but examples like (15) prove the contrary. Both for typical mass nouns as well as for SOMNs, there are multiple theories and possible explanations for their uncountability. For core mass nouns like *water* as well as granular aggregates like *sand*, Chierchia (2010, 116–118) argues that they are not countable because they have no stable minimal parts to distinguish, a prerequisite for counting operations: “The point is that there is no systematic basis for deciding which rice amounts qualify as rice atoms [i.e. minimal parts]” (Chierchia 2010, 118). However, he does not seem to have a satisfactory explanation for the uncountability of SOMNs: “The point is that tables and chairs are furniture, boots and shoes are footwear, and so on. Hence,

¹¹ This leads to the assumption that OMNs are only available in languages that distinguish between count and mass nouns (cf. Chierchia 2010, 111; Doetjes 2011, 2564); this question however has not been systematically analysed until now (but see Mihatsch/Kleineberg in prep. b for some preliminary analyses on Papiamentu).

¹² Related to this question of the uncountability of SOMNs is the theoretical interpretation of mass nouns in general in the framework of formal semantics. It is, however, outside the scope of this present study to fully account for all approaches to the distinction between mass and count nouns. It may be roughly summarised that authors differ in either assuming the mass-count distinction to be linked to ontology (cf. Link 1983; Chierchia 2010), or in considering it to be merely a grammatical distinction (cf. Krifka 1989; Rothstein 2010). There is a general tendency to consider mass nouns as root nouns and count nouns as being derived from them (cf. e.g. Borer 2005; Rothstein 2010); theories differ, however, as to the exact point and by what linguistic means this derivation is operated (see Rothstein 2017, chap. 4 for a comprehensive overview of these discussions).

it seems highly implausible that nouns like *furniture* are any vaguer than *table* or *chair*. It follows, that the mass like behaviour of nouns like *furniture* must come from a totally different source” (Chierchia 2010, 140).

Rothstein (2010a) also argues that we need a set of well-defined minimal parts which we can count, as for SOMNs she says that “the set of minimal elements is not lexically accessible and is not countable” (Rothstein 2010a, 362). For this reason, she distinguishes between naturally atomic entities like *boy* and *furniture*, which both refer to discrete individuals in the real world, and semantically atomic predicates, which linguistically code the referents as being discrete individuals. From a perspective of natural atomicity both *boy* and *furniture* are atomic, because they refer to discrete individuals in extra-linguistic reality (in contrast to *water*, for instance). From the point of view of semantic atomicity only *boy* is also semantically atomic. In this case, the linguistic label reflects the extra-linguistic features, but SOMNs like *furniture* are semantically non-atomic – that is, they are linguistically marked as having no atoms. A third explanation comes from Landman (2011). He also argues that we cannot count SOMNs (*neat mass nouns* in his terminology), because we do not know what to count: “[. . .] a plurality of *kitchenware*, like *the cup and saucer*, can count itself as *kitchenware*, and can also count as one” (Landman 2011, 34). Summarising these three approaches, we can cautiously conclude that we cannot count SOMNs, because the minimal parts of which they consist are linguistically blurred, so that we simply do not know what to count.

Because of their morphologically singular form, SOMNs, but also homogeneous mass nouns, are often classified as *singularia tantum* (cf. e.g. Wierzbicka 1984). SOMNs are, however, not exactly only morphologically singular and necessarily lacking an equivalent plural form, they are formally transnumeral. I borrow this notion from Biermann (1982, 230) who defines transnumeral nouns as being neither singular nor plural, but number neutral.¹³ For SOMNs, being neither compatible with the indefinite article (cf. *supra*), nor having an inflectional plural form, I classify them not as *singularia tantum*, but as transnumeral, neither singular nor plural. POMNs seem to be a little different: they are equally typically classified as *pluralia tantum*, they have no equivalent singular form and they are not compatible with the indefinite article. As has been briefly touched upon in

13 This notion of transnumerality is not defined uniformly. In addition to the form-oriented definition adopted in this present framework, there is also a meaning-oriented definition adopted by e.g. Greenberg (1977) and following him also Rijkhoff (2002, 45). They call those nouns transnumeral which do not specify whether they refer to a singularity or a plurality. This definition of transnumerality is thus similar to the notion of general number of Corbett (2000) (cf. *supra*).

the introduction, the classification as a plurale tantum however only refers to the formal features of a noun and not necessarily also to its semantics. There are nouns like *garden shears* that equally have no equivalent singular form and, which do not refer to a plurality, but to a single entity in the extra-linguistic reality (cf. Acquaviva 2008, 16–17). Moreover, there are so-called dense pluralia tantum like Fr. *épinards* ‘spinach’ that do refer to a plurality of entities, but which are so small and homogeneous that we do not construe them as discrete (cf. chap. 1.1). POMNs thus share, in contrast to a number of only formally plural nouns and also to SOMNs, with the inflectional plural the property of possible quantification. Lauwers (2014, 123–127; 2016, 274–275) attests for *pluriels collectifs hétérogènes* ‘collective heterogeneous plurals’ like Fr. *vivres* ‘food supplies’ or *ossements* ‘mortal remains/bones’ the possibility of being combined with distributive quantifiers like Fr. *divers* ‘various’ and, more restrictedly, also with numerals – these constructions are said to be rare, but possible. Fr. *gens* ‘people’, however, is said to be only compatible with distributive quantifiers like *plusieurs* or *quelques* ‘various/some’, but not with numerals, or at least this combination is said to be highly marked (cf. Grevisse/Goosse 2016, §510a). Because of the possibility of their combining with these kinds of quantifier, and maybe also with numerals, I assume for POMNs a lesser degree of unboundedness, as for SOMNs. The constituting entities of a POMN are easier to access because the outer frontier of the collection is permeable. In this respect, POMNs are similar to the inflectional plural. These aspects already indicate that there seem to be no clear-cut categories and kinds of collection noun, but rather a continuum, the properties of whose members are governed by various features. As has been shortly described, this has already been stated for different kinds of plurale tantum that may range from granular aggregates to heterogeneous collections (cf. Wierzbicka 1988, 499–560; Lauwers 2014; 2016).

In addition to the features of (non-)countability and of (non-)compatibility with certain determiners, there is another syntactic feature that defines the aspectual feature of (un-)boundedness of a noun. In chap. 1.1, I drew attention to the numerous parallels between verbal and nominal aspectuality and one of those parallels may relate to the nature of a noun’s external aspect. In his seminal paper, Krifka (1989) states for both mass nouns like Germ. *Apfelmus* ‘apple sauce’ and bare plurals like Germ. *Äpfel* ‘apples’ that they are not compatible with time-span adverbials like *in an hour*, but only with durative adverbials like *for an hour*. In contrast, singular object nouns like *Apfel* ‘apple’ or quantified nouns like *ein Pfund Apfelmus* ‘a pound of apple sauce’ as well as *drei Äpfel* ‘three apples’ are compatible with time-span-adverbials, but not with durative adverbials. These generalisations are exemplified in (16) and (17) (Krifka 1989, 227):

- (16) Bare plurals and mass nouns¹⁴
- a. Anna aß zehn Minuten lang Äpfel/Apfelmus.
'Anna ate apples/apple sauce for ten minutes.'
 - b. *Anna aß in zehn Minuten Äpfel/Apfelmus.
'Anna ate apples/apple sauce in ten minutes.'
- (17) Singular object nouns and quantified nouns
- a. *Anna aß zehn Minuten lang einen Apfel/drei Äpfel/ein Pfund Apfelmus.
'Anna ate an apple/three apples/a pound of apple sauce for ten minutes.'
 - b. Anna aß in zehn Minuten einen Apfel/drei Äpfel/ein Pfund Apfelmus.
'Anna ate an apple/three apples/a pound of apple sauce in ten minutes.'

The reason for these (in)compatibilities lies in the unboundedness of mass nouns and bare plurals and the boundedness of singular object nouns and quantified nouns. Unbounded entities refer cumulatively and therefore each part of this entity is equal to the entity, i.e. each portion of apple sauce is apple sauce. In contrast, this does not hold for quantised entities like *three apples*, as a part of *three apples* is unequal to *three apples* (but only one apple e.g.) (cf. Krifka 1989, 230–231). The same holds for telic and atelic predicates in the verbal domain. Each part of the atelic predicate of Germ. *laufen* 'to run' is like the predicate itself, whereas this is not true for telic predicates like Germ. *drei Kilometer laufen* 'to run three kilometres' (cf. Krifka 1989, 236). As a consequence, we can run for an hour, but not in an hour and we can run three kilometres in an hour, but not for an hour. These similarities in the verbal and the nominal domain are also reflected not only by homogeneous mass nouns like *apple sauce*, but also by OMNs like *furniture*. As indicated by Borer (2005, 121), OMNs behave like typical mass nouns and nouns denoting granular aggregates in not being compatible with time-span adverbials, but only with durative adverbials:

- (18) a. Kim ate meat for an hour/*in an hour.
b. Pat built furniture for two months/*in two months.
c. Robin shifted sand for half an hour/*in half an hour.

Her explanation is not based on cumulative reference but lies in the fact that OMNs like *furniture* do not denote a well-defined set of entities, but are vague in leaving open the question of how many and exactly which kinds of entity they comprise. The same reason why we cannot count OMNs thus also explains their

¹⁴ For more parallel characteristics of mass nouns and (bare) plurals see Lasersohn 2011.

incompatibility with time-span adverbials. If, however, we eliminate the vagueness of a mass noun, a granular aggregate like *sand* or an OMN like *furniture*, contexts like those in (19) become grammatically felicitous (Borer 2005, 121):

- (19) a. Kim ate more than enough meat in an hour.
 b. Pat built most furniture in two months.
 c. Robin shifted too much sand in half an hour.

Summarising this overview of the (un)boundedness of collection nouns, research done so far indicates a clear boundedness and therefore countability of typical CCNs like Fr. *équipe* ‘team’, a clear unboundedness and therefore uncountability of SOMNs like Sp. *ropa* ‘clothing’, and POMNs occupying the margin between boundedness and unboundedness, countability and uncountability. These aspects can be illustrated finally by the following example adapted from Moltmann (1997, 87). The Fr. POMN *fringues* ‘clothes’ in (20a) is felicitous in a context that implies countability, but the Fr. SOMN *vêtement* ‘clothing’ in (20b) is, however, weird in this context and to be able to enumerate a CCN like Fr. *tenue* ‘outfit’ (cf. (20c)) one needs the inflectional plural, as the outer boundedness blocks access to the constituting entities:

- (20) a. Jean énumère les fringues_{POMN}.
 ‘John enumerates the clothes.’
 b. ??Jean énumère le vêtement_{SOMN}.
 ‘John enumerates the clothing.’
 c. Jean énumère *la tenue_{CCN-SG}/les tenues_{CCN-PL}.
 ‘John enumerates *the outfit/the outfits.’

3.1.2 Internal plurality of count collective nouns and object mass nouns

Crucial to the definition of every type of collection noun is the referential internal plurality of discrete entities – irrespective of syntactic characteristics. In chap. 1.1, I summarised the various approaches of nominal aspectuality with respect to internal discreteness from a theoretical point of view, and I will now look at these theoretical aspects in more detail by applying them to CCNs and OMNs. We already saw that, on the one hand, the aspect of being comprised of discrete entities is not a dichotomous feature, but should rather be viewed as a continuum and that, on the other hand, various aspects like the heterogeneity of constituting entities can influence the linguistic characteristics reflected by them. In this respect, I already distinguished between rather homogeneous CCNs like *bunch*

and more heterogeneous ones like *family* or *team*. I also indicated that OMNs may differ in comprising either more homogeneous entities like *fruit* or more heterogeneous objects as in the case of *furniture*. I will now address the basic issue of the accessibility of this internal plurality and the question of how the constituting entities are structured within the collection.

A first approach to accessing the internal plurality of collection nouns is not necessarily a mere linguistic one, but a psycholinguistic one. Particularly with respect to SOMNs, research until now has mainly focused on the mismatch between the non-countability of these kinds of collection noun and, the nonetheless present internal plurality of them. Whereas SOMNs were still treated as ordinary mass nouns by Quine (1960, 91), who mentions examples like *footwear* along with *water*, more recent psycholinguistic studies focus on their underlying conceptual structure. In this respect, Barner/Snedeker (2005) examine the question of whether nouns like *furniture* can quantify over individuals with the help of quantity comparison and whether there are differences in this interpretation in language development. To analyse these questions, they showed adults and children alike pictures of substances like toothpaste as well as of discrete individuals like shoes or knives and forks. The substances were labelled with the respective mass noun, referents like the shoes were labelled with a plural object noun and those like the knives and forks with a SOMN, in this case *silverware*. In each picture, there was a character with two instances of the referent, the first was greater in volume and the second was more in number. The test subjects then had to answer the question “who has more silverware/more shoes/more toothpaste?” The results reveal that both adults and children interpret the quantity of SOMNs via the single individuals and not the overall volume. Consequently, they detect a significant difference between the interpretation of homogeneous mass nouns and SOMNs, but no difference between the interpretation of the latter and plural object nouns. The authors also found no differences between homogeneous or heterogeneous sets of objects representing the SOMN. That is to say, they obtained the same result with three forks and one larger fork and three sets of knife and fork and one large instance of this set (cf. Barner/Snedeker 2005, 52–53). The experiments of Barner/Snedeker (2005) prove that SOMNs quantify over individuals, although they are not directly countable. The interpretation of the quantity of a SOMN may, however, also be based on volume, given the right context. Rothstein (2017, 122–123) gives examples like the one in (21), where quantity evaluation based on volume is unproblematic because the context triggers it:

- (21) John got sick because he ate more fruit than Mary. She ate two apples and three strawberries. He ate a whole watermelon.

Her mostly theoretical impression is proven empirically for Brazilian Portuguese by Beviláqua/Pires de Oliveira (2014) who find that a mass context may trigger either a volume or a cardinal interpretation of SOMNs like Pt. *mobília* ‘furniture’ or *bagagem* ‘luggage’. The contexts in their tests mainly comprised loading objects into a truck or a car, for example for a move or a holiday trip.

A basic linguistic construction for testing the existence of some underlying structure of discrete entities of a noun is the combination with a so-called *stubbornly distributive predicate* (cf. Schwarzschild 2011).¹⁵ These kinds of predicate exclusively modify clearly distinguishable entities, and if there are none, the construction becomes infelicitous. Consequently, mass nouns like *wine* and nouns denoting granular aggregates like *sand* or *rice* are not, or at least hardly, compatible with adjectives like *round*, *small* or *long*, because their constituting minimal parts are conceived of as being a more or less homogeneous mass (cf. *supra*). OMNs like *furniture* however are absolutely fine in these constructions just as bare plurals, like *boxes*:

- (22) a. ?The wine is big/?The snow is round. (Schwarzschild 2011, 665)
 b. ?large rice (in the sense of long-grained rice)/??large sand (McCawley 1975, 319)
 c. ?The rice is long/round. (Hana Filip, personal communication, January 2020)
 d. The boxes are round. (Schwarzschild 2011, 664)
 e. The furniture in that nightclub is round./The mail in that bin is square and small. (Schwarzschild 2011, 670)

With CCNs, although these also refer to discrete entities, the combination of a stubbornly distributive predicate with such a noun is also infelicitous. We already saw in chap. 3.1.1 that the set profiling of nouns like Fr. *jury* ‘jury’ hinders access to the constituting entities (Borillo 1997, 113). Furthermore, adjectives like Fr. *grand* ‘big’, which allow both a collective and a distributive reading (‘numerous’ vs. ‘tall’), only create a collective interpretation in combination with CCNs: “un grand comité est un comité nombreux” (‘a big committee is a committee with many members [and not with tall members]’) (Flaux 1999, 474). The same holds for the example of Sp. *gentío grande* ‘big crowd’, already mentioned in chap. 3.1.1 (Bosque 1999, 54). With an OMN however, distributive readings of adjectives like Sp. *grande* ‘big’ are indeed possible: Sp. *gente grande* only has the possible interpretation of

¹⁵ A first marginal reference to this kind of adjective and their connection to the mass-count distinction was already made by Bunt (1979, 269).

‘tall people’ and not ‘many people’ (cf. Real Academia Española/Asociación de Academias de la Lengua Española 2009, 807). These considerations lead to the conclusion that: “die Indizien für Pluralität und die Indizien für Gegenstandscharakter [können] meist nicht gleichzeitig auftreten” (‘the indications of plurality and the indications of outer boundedness usually cannot occur simultaneously’) (Mihatsch 2006, 101). I have not come across any data on POMNs in combination with stubbornly distributive predicates, but my intuition tells me that they go equally well with these adjectives: Germ. *große Leute* ‘tall people’ for example only has the possible interpretation of people being tall and not of a large quantity of people. An exception to this generalisation may be those verbs that only imply plurality like *to gather* and which consequently are not compatible with e.g. singular object nouns. These verbs do not seem to be as restrictive as stubbornly distributive predicates, and constructions like Sp. *la familia se reunió* ‘the family got together’ are perfectly fine (cf. Bosque 1999, 44–45; Depraetere 2003, 102–103).

In contrast to this relatively clear picture of collection nouns in combination with stubbornly distributive predicates, other distributive predicates cross-linguistically show more heterogeneous features of compatibility with collection nouns. For English, both Chierchia (1998a, 86) and Rothstein (2010a, 379) find that neither OMNs like *furniture* (cf. (23a–b)) nor CCNs like *committee* (cf. (23c)) may be combined with reciprocals or other distributive predicates like *one after the other*. Only plural noun phrases are felicitous in these constructions (cf. (23d–e)):

- (23) a. *That furniture is leaning against each other.
 b. *The furniture is piled on top of each other.
 c. *Committee A fights each other.
 d. Those pieces of furniture are leaning against each other.
 e. Committee A and committee B fight each other.

This assumption, however, seems to be language-specific and, as assumed by Rothstein (2010a, 383–384), languages like Portuguese (both Brazilian and European) do accept this kind of construction, at least with mass collection nouns like Pt. *mobília* ‘furniture’ (cf. also Pires de Oliveira/Rothstein 2011a, 2157; 2011b, 237; Rothstein 2010b, 16, footnote 4; 2016a; 2016b):

- (24) *Mobília (dessa marca) encaixa uma na outra.* (Rothstein 2010a, 384)
 ‘Pieces of furniture (of this brand) fit into each other.’

There are two issues with this kind of cross-linguistic variation. On the one hand, Rothstein and Pires de Oliveira do not explain their findings, but simply say that there are cross-linguistic differences with respect to sensitivity to natural or seman-

tic atomicity (cf. Pires de Oliveira/Rothstein 2011b, 253). This is to say that both Engl. *furniture* and Pt. *mobília* are naturally atomic, but only *mobília* is also semantically atomic in accepting distributive predicates (cf. *supra* for the categorisation between natural and semantic atomicity). On the other hand, there is no clear empirical evidence for this finding, which is mostly based on introspection. Braga et al. (2010) examine in this context the general acceptability of Portuguese bare singulars in combination with more or less distributive predicates. The test sentences were structured according to the distinction made by Rothstein (2010a) between natural and semantic atomicity, with the researchers testing six mass nouns and six count nouns of which half were naturally atomic and the other half non-naturally atomic. Example (25) shows the three naturally atomic mass nouns, which are relevant for this present elaboration (cf. Braga et al. 2010, 88):

- (25) a. Cabelo se enrosca.
 ‘Hair curls up.’
 b. Faqueiro pesa doce quilos.
 ‘Cutlery weighs twelve kilos.’
 c. Móvel encosta uma na outra.
 ‘Furniture leans against each other.’

Each sentence was evaluated by 50 native speakers of Brazilian Portuguese. The results show that the sentences containing a naturally atomic mass noun were rated less acceptable than naturally atomic count nouns like *criança* ‘child’ or *cachorro* ‘dog’. The authors thus conclude that “speakers seem to be paying attention to the combination of natural and semantic atomicity” (Braga et al. 2010, 91). However, the results were not very clear and speakers seem to have had basic problems with combining the bare singular with distributive predicates, irrespective of atomicity, whether semantic or natural (cf. Braga et al. 2010, 92). Unfortunately, the topic does not seem to have been pursued any further by the authors, so at present there is only this inconclusive result. Against the backdrop of this present approach of nominal aspectuality, there is additionally the issue of a different aspectual configuration of the tested naturally atomic mass nouns: *faqueiro* ‘cutlery’ and *mobília* ‘furniture’ are constituted of heterogeneous, clearly distinguishable objects. Even if we accept, however, the categorisation that *cabelo* ‘hair’ is naturally atomic, it still consists of much smaller, more homogeneous entities. According to my argument in chap. 2, I would thus categorise *cabelo* as a granular aggregate and would consequently expect a different linguistic behaviour – most likely a lesser degree of acceptability with predicates accessing the constituting entities. Distributive predicates are consequently good candidates for indicating the internal plurality of collections, but so far there is still uncertainty as to the

degree to which cross-linguistic variation, saliency of the constituting members or boundedness play a role in the acceptability of this kind of construction.

Another possible indicator of the internal plurality of a collection is the so-called *constructio ad sensum* or synesis, i.e. grammatical agreement depending on semantics and not on syntax or morphology. Typical for these kinds of construction is a syntactic mismatch in number agreement between subject and predicate, but also gender may agree according to the meaning and not the form. A common example for the latter type is Germ. *das Mädchen . . . sie* ‘the girl . . . she’, where the subject *Mädchen* is of neuter gender, but the anaphorical pronoun *sie* of feminine gender, agreeing with the meaning of the noun and not with the grammatical information.¹⁶ Especially in British English, the phenomenon is generally common with human collective nouns like *committee* which can either occur with a verb in the singular (syntactic concord) or in the plural (semantic concord): *The committee is/are wrong*.¹⁷ Many authors treating English collective nouns take this possibility of plural agreement with singular collective nouns as a definitional criterion for this category (cf. DLP, s.v. *collective*; Visser 1963, 62; Nixon 1972; Quirk 1985, 316–317; cf. also Cruse 1986, 176).¹⁸ Although for Romance languages and also for German, this direct mismatch of grammatical agreement is considered impossible (cf. Dubois/Dubois Charlier 1996, 125; Flaux 1996–1997, 33; Borillo 1997, 111; Joosten 2010, 33, footnote 12; Gross 2011, 65):

- (26) a. *Meine Familie sind Frühaufsteher. (Wegerer 2012, 117)
 ‘My family are early birds.’
 b. ?L’armée ont froid aux pieds. (Lammert/Lecolle 2014, 212)
 ‘The army have cold feet.’
 c. *L’assemblée des fonctionnaires sont sortis de l’auditoire. (Michaux 1992, 113)
 ‘The assembly of officials have left the audience.’
 d. *La familia estaban de acuerdo. (Bosque 2000, 13)
 ‘The family agreed_{PL}.’

¹⁶ Interestingly, as shown by Braun/Haig (2010), the chance of semantic agreement grows with increasing age of the girl. This is in accordance with the agreement hierarchy of Corbett (1979), which predicts semantic agreement to be more likely with predicates and attributes than with pronouns. Consequently, the more salient the feature of being feminine gets, the more likely is agreement with attributes and thus semantic aspects rather than morphological ones (cf. also *infra*).

¹⁷ In languages like Georgian or Oromo (Cushitic, Kenya and Ethiopia), there is also the possibility for singular verbal agreement with a plurality of individuals, since in those languages “‘number’ agreement is always with the set [of individuals]” (Rijkhoff 2002, 105).

¹⁸ For restrictions on semantic agreement with collective nouns see Smith 2017.

So-called constructiones ad sensum become the more probable the greater the distance between the subject and its predicate (cf. Gili Gaya 1973, 31–32; Martínez 1999, 2767–2768). In this respect, examples like in (27) are perfectly fine, in contrast to e.g. (26d):

- (27) El público, al menos los convencidos, aplaudían calurosamente. (Martínez 1999, 2767)
 ‘The audience, at least the convinced people, applauded_{PL} warmly.’

In this context, Sánchez Avedaño (2007), analyses verbal agreement with the Spanish SOMN *gente* ‘people’. His examination shows that in cases of a direct connection between the collection noun and the respective predicate the agreement is grammatical in 93% of cases, i.e. the verb is in the singular. In contrast, when there is another sentence between the two elements, he finds constructiones ad sensum in 89% of cases, i.e. with a verb in the plural (cf. also Levin 2001, 92–99 for a similar analysis on English collective nouns). Millán Orozco (1977, 142) moreover finds that, at least in his example of Mexican Spanish, there are only simple nouns which allow constructiones ad sensum – derivatives like Sp. *personal* ‘staff’ do not seem to permit these kinds of syntactic mismatch. It should also be noted that semantic agreement with collection nouns is in general mostly restricted to oral communication (cf. Flaux 1996–1997, 35; Bosque 1999, 38; Levin 2001, 120–121; Depraetere 2003, 112). As shown by Joosten et al. (2007) the probability of semantic or syntactic agreement may depend on the degree of set profiling of a noun. As in the case of combination with stubbornly distributive predicates, plural agreement seems to be thereby facilitated with OMNs in contrast to CCNs (cf. Ortega/Morera 1981–1982, 650–651). For instance, the Sp. SOMN *gente* ‘people’ may be combined with a verb in the plural, but the Sp. CCN *familia* ‘family’ may not. Note however, that the examples given by Ortega/Morera (1981–1982, 650–651) do not refer to direct combination of the singular subject and the plural predicate, but to a construction with space in between these two syntactic elements (cf. (28a)). Furthermore, the counter example of a CCN is inanimate in contrast to the OMN (cf. (28b)), another aspect which may influence synesis (cf. *infra*). At least for German, constructiones ad sensum do not seem possible with either OMNs or CCNs.¹⁹

¹⁹ There is one systematic study on cases of doubt concerning verbal agreement in German. With respect to synesis with collective nouns, Wegerer (2012, 278–283) tested copula constructions with plural predicates like *Mein Lieblingsobst ist/sind Orangen* (‘My favourite fruit is/are oranges’), where verbal agreement does not necessarily depend on the type of collection noun, but rather on whether *Obst* (singular) or *Orangen* (plural) is interpreted as the subject.

- (28) a. La gente que acabó no saben leer ni escribir.
 ‘The people who finished can’t read or write.’
 b. El enjambre desapareció entre los árboles.
 ‘The swarm vanished into the woods.’

In quantifying binominal constructions, it is additionally possible to focus either on the grammatical or on the semantic nucleus of the complex construction. In the former case, agreement would be in the singular, as shown by (29a), and in the latter case agreement would be in the plural, as exemplified by (29b):²⁰

- (29) a. Une horde de vigiles faisait office de comité d’accueil. (e-mail, cit. in: Lammert/Lecolle 2014, 212)
 ‘A horde of guards acted_{SG} as a welcoming committee.’
 b. Une compagnie d’oiseaux tourbillonnaient dans le ciel bleu. (G. Flaubert, cit. in: Lammert/Lecolle 2014, 212)
 ‘A swarm of birds whirled_{PL} in the blue sky.’

A final indicator for the internal plurality of a collection are pronominal anaphors which may be in the plural. Following the *agreement hierarchy* of Corbett (1979), semantic agreement with personal pronouns is more likely to occur than a constructio ad sensum with verbal semantic agreement (cf. also Nixon 1972, 123; Corbett 2000, 189–190; Joosten et al. 2007, 97–101; Soler Arechalde 2008, 1143–1145). This is illustrated in (30) where the predicate agrees syntactically with the collection noun, but the anaphoric personal pronoun agrees semantically:

- (30) a. J’ai de la famille en Normandie. Ils_{PL} habitent au bord de la mer. (Flaux 1999, 482)
 ‘I have family in Normandy. They live by the sea.’
 b. L’équipe est_{SG} sur le point de gagner – ils_{PL} se surpassent tous pour ce match décisif. (Lammert/Lecolle 2014, 212)
 ‘The team is about to win – they’re all outdoing each other for this decisive match.’

20 In these quantifying binominal constructions, it is possible for a collective noun to be reanalysed as a pseudo-quantifier. For instance, Sp. *hatajo* ‘small herd’ may function as a quantifier in constructions like *un hatajo de arribistas* ‘a bunch of careerists’ (and not ‘a small herd of careerists’). The more it becomes a quantifier the more it loses its status as the syntactic head of the binominal construction and the more verbal agreement in the plural becomes plausible (cf. Verwecken 2012; 2015; Verwecken/Cornillie 2012; Verwecken/Delbecque 2015; Delbecque/Verwecken 2014; for French cf. also Borillo 1997, 108–109; Flaux 1999, 473–474; Lammert/Lecolle 2014, 214).

- c. Todo el alumnado pensaba_{SG} que los profesores se ocupan poco de ellos_{PL}. (Bosque 2000, 15)
 ‘All the pupils thought that the teachers care little for them.’

One may already have noted that all the examples of collection noun given in these last few paragraphs mostly comprise human entities. And there is indeed the tendency in all these constructions that they are only possible with human collections. The underlying principle of this generalisation can be explained by the animacy hierarchy introduced in chap. 1.2.1. The reasons why human beings seem to be more salient to our own perception and consequently may be treated differently by grammar are numerous. In addition to their perceptible individuality, factors such as agency (humans are more likely to assume the role of an agent than other animate or inanimate entities), empathy (as belonging to our own species, other human beings are more relevant to us) or the possibility of being in motion may also play a part (see Vihman/Nelson 2019 for an excellent up-to-date overview of the role of animacy at the interface between grammar and cognition). With respect to collection nouns and the grammatical indication of their internal plurality, it could be that human entities, as a result of their saliency, facilitate linguistic accessibility to them via plural verbal agreement or plural personal pronouns. The studies of Joosten et al. (2007) additionally suggest that the number of people making up the collection as well as the process of formation and the level of involvement of the members may also be factors influencing entity accessibility of collections. Semantic agreement is not or only marginally possible with animate or inanimate entities, as shown by (31a) and (31b). This is however also true for collection nouns that massify the constituting human individuals in such a way that they are not accessible anymore. This is shown by (31c), where Fr. *foule* ‘crowd’ implies an indefinite, unordered mass of people.

- (31) a. *Dans mon troupeau, ils sont tous malades. (Lammert 2010, 83)
 ‘In my herd, they are all ill.’
 b. *Cet essaim ne m’a pas épargné: ils m’ont tous attaqué. (Lammert 2010, 83)
 ‘This swarm has not spared me: They all attacked me.’
 c. *J’étais pris dans la foule. Ils ont failli m’étouffer. (Kleiber 1994, 174)
 ‘I was caught in the crowd. They did not suffocate me.’

In addition to differences with respect to the various linguistic operations of accessing the plurality of the constituting entities of a collection noun, these entities may also be structured in various ways, standing in different relations to their superordinate noun that designates the collection itself. Following Cruse

(1986), I differentiate between two types of branching hierarchical lexical relation: hyponymy and meronymy. Both principles describe the relation between a subordinate element and a dominating superordinate element, whereby their branching character additionally determines various parallel levels in the hierarchy. These kinds of lexical hierarchy are typically characterised by two properties: asymmetry and chaining. The relations are asymmetric in being directional: a superordinate element A may stand in a relation R to a subordinate element B, but B cannot stand in the same relation R to A. The relations are catenary in theoretically going on eternally as in, for example, A is the father of B, B is the father of C, C is the father of D and so on (cf. Cruse 1986, 112–114). Those properties are typical for hyponymies and meronymies; both are hierarchical, branching, asymmetric and chained lexical relations. The notion of hyponymy, first introduced to linguistic theory by Lyons (1972), is typically expressed by the phrase *X is a type of Y* and implies class inclusion. In this respect, a sparrow is a kind of bird and a bird is a kind of animal (but not vice versa). The branching character of typical hyponymies like this is represented by other co-hyponyms like *dog*, which also stand in a kind-of relation to the hyperonym *animal*, but which are mutually exclusive to other co-hyponyms. Meronymy, on the other hand, is typically expressed by the construction *X is a part of Y*, but “instead of there being a single clearly distinguished relation, there is in reality a numerous family of more-or-less similar relations” (Cruse 1986, 160). In this respect, Winston/Chaffin/Herrmann (1987) differentiate six types of meronymy including e.g. component–integral object, member–collection or feature–activity. All these relations have the underlying part-whole definition in common, whereby a handle is part of a cup, a card is part of a deck and paying is part of the activity of shopping, but they differ in, for instance, being separable or not from the whole. We can easily separate the handle of the cup and it still remains a cup, but paying is an essential part of shopping and cannot be removed without changing the property of the whole (otherwise it would be stealing). As in the case of hyponymy, meronymy also shows the typical properties of asymmetry, branching and chaining.

In addition to this hierarchical structuring of the level of language, lexical relations seem also to play a role in cognition. Although natural entities like animals show various indications of conceptual autonomy (cf. Rosch et al. 1976), word association tests and linguistic disorders like aphasia give rise to the assumption that the mental lexicon is structured by these kinds of relation. According to Aitchison (2012, 101–102) the links most often found in word association tests can be pinned down to coordination (words on the same level of detail), collocation (words occurring together in discourse), superordination (hyperonyms of the word named) and synonymy (meaning similarity). Coordination is found most often and synonymy the least. In addition, aphasic patients often mix up coordi-

nates like *lemon* and *orange* both in production and comprehension (cf. Aitchison 2012, 103–104). These tests indicate that the mental lexicon is not made up of simple word lists, but that it is very likely structured by intertwined networks determined by various conceptual links. It is assumed in this respect that syntagmatic relations like collocation are cognitively primary in opposition to paradigmatic relations like coordination or synonymy. That is to say that children at kindergarten age give more collocational answers in word association tests than adults do. Only when they go to school do paradigmatic relations become more relevant and are named more in these kinds of tests. A small child would say e.g. *crayon* when hearing the word *black*, an adult would in contrast say something like *white*, i.e. the antonym, a paradigmatically related word. The change from one to the other kind of relating words is called the *paradigmatic shift*, most likely caused by school socialisation (cf. Murphy 2000, 340–341). There is thus evidence for some kinds of semantic relation in the mental lexicon, but how exactly they come into being, how they change and whether these relations are between concepts or words are still questions awaiting further research (cf. also Mihatsch 2006, 27). Important for this present framework is the fact that the lexical relations expressed by linguistic constructions seem not only to be based on mere language-specific semantic aspects, they also seem to have some kind of conceptual representation. In this respect, it is assumed that hyponymy is based on perceptual similarity and meronymy on perceptual contiguity (cf. Mihatsch 2006, 29–31).

After this short introduction to lexical hierarchies, differences in the structuring of the constituting entities in the denotation of various collection nouns can now be elaborated. First, CCNs are generally said to be constituted through a meronymic collection–member relation. With respect to this meronymic subtype, Lecolle (1998, 48–49) has determined three main properties of CCNs. Firstly, the entities which make up the collection must be of the same category and they have to represent themselves as wholes (cf. also Borillo 1997). This property is the prerequisite for fusion and division of CCNs, as already described in chap. 3.1.1 (cf. Flaux 1999, 475–477). A second consequence of the categorical homogeneity of the constituting members of a collection is the possibility of partitive constructions of the kind <N1_{collection} of N2_{members}>, and some authors in fact take this property as a defining aspect of *true* CCNs (cf. e.g. Barker 1992, 70; Michaux 1992, 111). Typical examples for this kind of construction are e.g. Fr. *un bouquet de fleurs* ‘a bouquet of flowers’, *un troupeau de moutons* ‘a herd of sheep’ or *un trousseau de clefs* ‘a bunch of keys’ (cf. Borillo 1997, 107; Lecolle 1998, 59–60; Flaux 1999, 474). This property distinguishes, for instance, the collection–member relation from other meronymic relations, as they are not possible in these constructions since they have no homogeneous members: Fr. **un vélo de pédales* ‘a bicycle of pedals’ (cf. Lecolle 1998, 60). Not all kinds of CCN may, however, take part in those partitive

constructions. For collection nouns, which are already semantically saturated, an additional mention of the members of the collection may be redundant, as is the case of Fr. *?une chorale de choristes* ‘a choir of choristers’. The redundancy may be resolved if the constituting members are further specified, as in Fr. *une chorale de seize choristes* ‘a choir of sixteen choristers’ (cf. Borillo 1997, 107; Lecolle 1998, 59–60; Lammert/Lecolle 2014, 213). These examples can be explained by the fact that, following Mihatsch (2006, 102), a decreasing syntagmaticity of a collection noun correlates with an increasing specificity of the referential members of the collection. Sp. *rebaño* ‘flock of sheep’ has a low syntagmaticity, because the constituting elements are already specified by the collection itself. Examples like Fr. *groupe* ‘group’ or *multitude* ‘multitude’ in contrast do not specify their members in advance and therefore need a prepositional complement to be fully functional (cf. also Flaax 1998, 176; see also *supra*). Second, the reference for defining the collection is represented by the plurality of entities and not the collection itself. The third property of CCNs, according to Lecolle (1998), is that the collection can simply be defined by a list of its constituting members. The two latter properties imply that collections of this type are defined *bottom-up*, i.e. taking the constituting members as a starting point and not the whole. As shown by Ising (2019, 53–54), the collection and the plurality constituting it are referentially identical. Following from this, together with the fundamentally independent character of the members with respect to the whole, entails that typical meronymy tests like *X has a Y/X has Ys* (cf. Cruse 1986, 160) do not apply to CCNs (cf. (32)). Tests which however take the members as a reference point are absolutely fine (cf. (33)) (Lecolle 1998, 50; 52 f.; cf. also Cruse 1986, 176; Borillo 1997, 107; Joosten 2010, 30).

- (32) a. ?Un archipel (a/comporte/comprend) des îles.
 ‘An archipelago (has/comprises) islands.’
 b. ?Un bouquet (a/comporte/comprend) des fleurs.
 ‘A bouquet (has/comprises) flowers.’
- (33) a. Un archipel (se compose/est composé) d’îles.
 ‘An archipelago (consists/is composed) of islands.’
 b. Un bouquet (se compose/est composé) de fleurs.
 ‘A bouquet (consists/is composed) of flowers.’

For being meronymic, the relation that holds together the plurality of entities of a CCN is of a contiguous nature. The exact type of contiguity may however vary and “spatio-temporal (*archipel, troupeau*), social (*couple, tribu*), cooperative (*club, armée*), or functional (*couvert, paire* (de chaussures))” (Joosten 2010, 38) relations are possible. Summarising the aspects regarding the internal structure

of CCNs, their constituting plurality may not be directly accessible through linguistic operations, but it can still be categorised by various features. Firstly, the relation between the constituting members and the collection is a meronymic collection–member relation based on various types of contiguity. Secondly, the constituting entities of a typical CCN are of the same ontological category. Thirdly, CCNs mostly take their members as a reference point and not the collection.

The second type of collection noun I distinguish in the framework of this present study are SOMNs. Whereas research generally agrees on the exclusively meronymic properties of typical CCNs, the relation between a SOMN and its subordinate constituting entities seems to be characterised by various lexical hierarchies:

These [SOMNs] are on the fuzzy boundary between classes and collections involving both similarity and spatial proximity. As a result we can say that, for example, a chair is both a kind of furniture (class inclusion) and an item of furniture (member-collection).

(Winston/Chaffin/Herrmann 1987, 428)

A *bra*, for instance, is not only *part* of underwear: it is also a *kind* of underwear.

(Joosten 2006, 79)

These two observations may be absolutely true, but SOMNs are neither good hyperonyms nor good holonyms. A collection noun like *furniture* or *jewellery* is not a good candidate for a typical superordinate given the fact that its constituting members are not similar from an intensional or perceptual point of view. A sofa, a cupboard and a table are not similar in sharing a portion of the same semantic features, neither do they necessarily share perceptual similarities. However, pieces of furniture, of jewellery or of clothing respectively share the property of having the same function (cf. Wierzbicka 1984, 320–321; 1988, 512–514). Consequently, there is nothing wrong in saying that a chair is a kind of furniture, but being furniture is not necessarily a crucial definitional feature for a chair, but rather being a kind of seat. For these reasons, Cruse (1986, 97) calls SOMNs like *cutlery* quasi-superordinates. The restriction to functional similarity seems mostly to apply to artefactual collections such as the those cited above. Animate collection nouns like *poultry* seem to be better suited for a more typical hyponymic relation between the constituting entities and the noun denoting the collection (cf. Joosten 2006, 78–79; 2010, 38–39). The heterogeneity of the constituting members of a SOMN, which is necessary for their functionality, also hinders them from being part of a typical meronymic collection–member relation. As described above, the prerequisite for many features in terms of this relation is ontological homogeneity. As a consequence, SOMNs are e.g. not compatible with binominal partitive constructions and SOMNs cannot determine their constituting entities by means of intensional implications. With respect to the former, it is not possible

to say Fr. **meuble des chaises anciennes* ‘furniture of antiques chairs’, because the SOMN requires heterogeneous members (cf. Flaux 1999, 484), just like Fr. *vélo* (cf. *supra*). In opposition to this, SOMNs, at least those referring to artefacts, need a contiguity relation between their constituting entities: to count as furniture, the table, sofa and cupboard have to be in the same room, otherwise they would be seats or simply artefacts made of wood. Wierzbicka (1984, 321) summarises this as the requirement for “unity of place”.

These aspects are supported by a number of psycholinguistic studies. I have already mentioned the quantity comparison experiments of Barner/Snedeker (2005) who found evidence for the existence of some conceptual structure of more than one individual in the context of SOMNs. However, they did not find any differences between the quantity judgements of a homogeneous or a more heterogeneous set of entities representing this SOMN. Grimm/Levin (2012), however, find that a greater heterogeneity of referents implying a greater functionality may dominate a greater cardinality of less functional referents, supporting the “similarity of function” introduced by Wierzbicka (1984, 320–321). They introduced participants in their experiments to a certain context, like friend A and B who want to furnish their rooms. Friend A has fewer entities, having a higher functionality because they are more heterogeneous, friend B has more entities having a lesser functionality because they are more homogeneous – e.g. a sofa, an easy chair, a coffee table and a bookcase (4 items) vs. one table and four chairs (5 items). At least for some of the tested SOMNs (*furniture*, *jewellery* and *change*), participants rated the quantity of the fewer pieces as more, because of their higher functionality. This was, however, not the case for *luggage*, *mail* and *ammunition*. The effect was only found when a certain context was introduced to the participants and neutralised when the context was absent.

The ambiguous character of SOMNs with regard to the lexical hierarchies they are part of is supported by studies from Markman (1985) as well as by Wisniewski/Imai/Casey (1996). Markman (1985) asks why some superordinate categories are mass in the world’s languages and hypothesises that mass superordinates like *furniture* help children learn these categories. To examine this hypothesis, the author taught 4-year-old children new categories and labelled the categories with a nonsense word, either in mass or in count syntax. For example, the four objects of soap, shampoo, a comb and toothpaste are either labelled with “this is veb” (mass) or “these are veb” (count). The results show that children are better at correctly retrieving the mass category than the count category when e.g. they are later asked to label the objects themselves. Markman (1985, 51) proposes that the mass character of these category-denoting nouns helps children to learn class-inclusion relations, because it “simplifies the hierarchical representation [yet remaining] faithful to the inclusion relation.” The reason for this lies

in the collective character of SOMNs. They display relations both of contiguity and similarity, where the latter is necessary for categorising, while the former is more primitive and ontologically acquired earlier (cf. Markman/Horton/McLanahan 1980; Callanan/Markman 1982). Wisniewski/Imai/Casey (1996) furthermore asked adults to evaluate whether objects of count and mass categories like musical instruments and jewellery are either more similar or occur more often together, to examine the proposition of Wierzbicka (1988) that SOMNs are constituted by a contiguity relation. The authors do indeed find that participants judged members of SOMNs to occur more often together than those of count categories. They did not, however, find any differences between count and mass categories with regard to the similarity relation. Whereas SOMNs like *furniture* are consequently characterised by both contiguity and similarity, entities of typical count superordinates like *vehicle* are judged as being only similar.

In conclusion, SOMNs indeed lie on the fuzzy boundary between the two lexical hierarchies of class inclusion and member-collection. The state of the art as summarised here consequently suggests that they may function as superordinates of both hyponymies and meronymies but, at the same time, they do not represent typical candidates for this position. The examples given in the literature additionally suggest that there seem to be differences between artefact and animate/inanimate SOMNs.

For the third type of collection noun, POMNs, there seems to be no consensus about the kind or kinds of lexical hierarchy that link the POMN and its constituting members. Wierzbicka (1984) is convinced that POMNs like *refreshments*, *clothes*, *belongings* or *groceries* are primarily constituted by the principle of contiguity: “All the members of the class in question designate heterogeneous collections of things, things which at some time are all in one place for the same reason” (Wierzbicka 1984, 321).²¹ With the help of this argument, she also explains the difference between *clothes* and *clothing*. Whereas *clothes* denotes “groups of things which could jointly ‘clothe’ one body, i.e. which could occur in the same place (on one person’s body), at one time” (Wierzbicka 1985, 283), its SOMN counterpart *clothing* may refer to “any miscellaneous collection of items because it is not thought of in terms of groups of things to be used jointly, at the same time, on the same body” (Wierzbicka 1985, 283). Put differently, POMNs do not seem to have any stable extension, but only seem to update their constituting members every time anew in discourse. In this respect, they resemble ad hoc

²¹ Related to this may also be the inherent ambiguity of the inflectional plural which may either have a distributive or a collective interpretation (cf. Lasersohn 2011, 1147–1150). The possibility of this latter interpretation may then be a semantic link to POMNs which denote mostly context-dependent collections (I heartily thank W. Mihatsch for this valuable comment).

categories like “things to take from one’s home during a fire”. In both cases, the collection of objects is not necessarily characterised by the feature of similarity, it is rather the common purpose of the different entities making up the collection. There is consequently the possibility of these kinds of POMN originating in ad hoc categories (cf. Barsalou 1983; Wisniewski/Clancy/Tillman 2005, 121). This would suggest that POMNs are more characterised by being part of a meronymic hierarchical structure (milk and eggs are part of groceries) than of a hyponymic structure – at least from an intensional point of view. In this respect, Wierzbicka (1984, 322) sums up that “the word *eggs* cannot be defined as a kind of groceries.” In contrast to this, Lauwers (2014, 125–126) assumes that POMNs like Fr. *vivres* ‘food supplies’ or *viscères* ‘intestines’ are indeed able to function as hyperonyms, provided that the prerequisite of heterogeneity is given (cf. also Lammert 2016 on the same impression on Fr. *fringues* ‘clothes’ and *excréments* ‘excrements’):

- (34) Les pommes de terre, le lait et le sucre sont des vivres.
 ‘Potatoes, milk and sugar are food supplies.’

Summing up this brief overview of the status of POMNs with regard to the hierarchical relations they form part of – and indeed research on this topic is scant – I must simply conclude that it is not clear. This may also arise from the fact that different authors have considered a highly heterogeneous class of nouns – animates and inanimates, natural entities and artefacts. It may be that the type of referent making up the collection labelled by the POMN influences the exact nature of the hierarchical relation between subordinates and superordinates – we have already seen this with regard to SOMNs (cf. *supra*).

In contrast to non-human OMNs, human OMNs – singular and plural – seem to be mostly part of meronymic hierarchies and do not accept a status as hyperonym. In this respect, Fr. *gens* and Germ. *Leute* ‘people’ are both spatio-temporally bounded and do not imply any stable semantic characteristics which are the prerequisite for similarity-based hyponymic relations. In contrast, the exact nature of their referents has always to be specified in the context, where they denote an undefined, but still contiguous collection of human beings (cf. Mihatsch 2015b for empirical evidence; cf. also Lang 2000).

In addition to the aspects discussed in chap. 3.1.1, this overview of the state of the art on the linguistic accessibility of the internal plurality of different types of collection noun revealed further differences between them. The necessary definitional criterion of collection nouns, their internal plurality of discrete objects, is fulfilled by all of these nominal types, irrespective of e.g. their countability features. The present discussion of the research done so far, however, gave rise to the assumption of a continuum of accessibility of this internal plurality. This accessi-

bility depends on various factors, which are not only limited to the mere external boundedness of collections, but additional aspects such as animacy also come into play. Moreover, collections show differences in the exact structure of their constituting entities, whether dominantly meronymic, rather hyponymic or both. Especially in the case of POMNs, many issues like e.g. the hierarchical structure of which they form a part, are still unclear. These issues will be addressed in chap. 4 where I present my own empirical research on this topic.

3.1.3 An integral semantic-syntactic model of collections

Based on the review of the state of the art in chaps. 3.1.1 and 3.1.2, it is now possible to create a first integral semantic-syntactic model of collection nouns taking into consideration not only the three different types discussed so far, but also related aspects such as the categorical type of entities making up the collection. The following model sums up the aspects on mostly English and French collection nouns as these are the most represented languages in research on this topic so far. However, keeping in mind the discussion in chap. 1, I consider the Germanic as well as Romance languages as being basically of one aspectual type, namely number marking languages, which may differ in their actual expression of nominal aspectuality (cf. e.g. Table 1.2) but in general behave in the same way. Recapitulating the remarks about different nominal types with regard to their aspectual coding, one may cautiously claim that the model described in the following represents on the one hand the semantic-syntactic features of collection nouns as described in the last two chapters, and, on the other, the underlying conceptual principles reflected by language.

As already indicated in chaps. 1.1 and 2, I consider only clearly distinguishable entities in this present work; as a consequence, granular aggregates as well as substances are not considered. With respect to these names for objects, I referred to the aspectual typology of Jackendoff (1991a, 20) in (3), repeated as (35) for convenience:

- (35) + bounded, – internal structure: individuals (a pig)
 + bounded, + internal structure: groups (a committee)
 – bounded, + internal structure: plural individuals, OMNs²² (buses, cattle)

²² The original representation refers to unbounded entities having an internal structure as *aggregates* (cf. (3)). I have changed this denomination here because the terminological choice is a little misleading in this framework, since I refer to aggregates only in the case of grains or granules like rice or sand (cf. chap. 2).

After the discussion of the different issues regarding various types of collection, one has to admit that the mere 2×2 matrix of aspectual types does not suffice to account for all the nuances and tendencies described in chaps. 3.1.1 and 3.1.2. OMNs in general as well as plural individuals may all be unbounded with an internal structure, but this does not take into consideration the different levels of accessibility of this internal structure. Whereas plural individuals are easily compatible with numerals and distributive quantifiers, we saw that POMNs are possibly only combinable with the latter and SOMNs are not compatible with any distributive quantifier at all, including the indefinite article. To account for this gradualness, I refine the dichotomous structure of Jackendoff. An initial refinement is done with the aid of the continuous typology of internal aspect following Meisterfeld (1998, 44–52). He distinguishes between various nominal types of internal discreteness on the basis of their respective levels of cohesion:

bare plurals < pluralia tantum < amalgams (CCNs/SOMNs) < aggregates
< (homogeneous) mass nouns

The first pole of this continuum is represented by bare plurals like Germ. *Häuser* ‘houses’, where nothing holds together the individuals and where there is a discrete juxtaposition of various individual objects. Homogeneous mass nouns like Germ. *Wasser* ‘water’ represent the other pole: here the level of cohesion is so high that we can no longer discern any discrete objects. Between these two poles are pluralia tantum like Germ. *Worte* which show a higher degree of cohesion than the bare plural *Wörter*. In English, both nouns can be translated as ‘words’, but *Worte* generally refers to words spoken with a special purpose in a specific context, and therefore shows a higher degree of cohesion than *Wörter* which just refers to various words. Meisterfeld (1998, 50–51) does not differentiate between countable or uncountable collection nouns in the morphological singular, but refers to both as amalgams, in terms of a complete connection of various elements. As exemplars for this category, he names Germ. *Gebäck* ‘biscuits’, *Gebälk* ‘timberwork’ or *Zaumzeug* ‘bridle’, all SOMNs, but also mentions various French collective suffixes that can derive exemplars of this category. Suffixes like *-erie*, *-age* or *-ure*, however, are able to derive both CCNs and SOMNs (cf. also chap. 3.2). Amalgams show an even higher degree of cohesion than pluralia tantum because they are constituted by (spatial) contiguity, as also shown in chap. 3.1.2. Aggregates, but also abstract nouns like *linguistics*, are finally very near to the pole of homogeneous mass nouns, and only differ in the qualities of possible distinct density, surface structure and so forth. Following the considerations of the last two chapters, I expand this continuum based on cohesion to a continuum of the accessibility of internal plurality which is not only based on cohesion, but also saliency of the constituting members.

A second refinement of the typology of Jackendoff (1991a) regards the dichotomous feature of boundedness. As shown by various authors examining countability (cf. i.a. Allan 1980; Grimm 2018) and also my own considerations of chaps. 1.2, 3.1.1 and 3.1.2, countability and the underlying aspectual feature of boundedness seem not to be binary, but again rather a continuum. Countability preferences of certain nouns may depend on the construal of the constituting entities they denote, but also on their use in discourse. As shown convincingly by a combination of corpus and acceptability studies on French and Dutch carried about by Vermote/Lauwers/Cuyper (2017), the acceptability of mass noun uses of fruit-denoting nouns increases with their usage frequency. In other words, the more a noun occurs in certain syntactic contexts, the more it becomes acceptable as such. This is in line with recent usage-based approaches in cognitive linguistics, also adopted in the framework of this present study (cf. in particular chap. 5). In addition, some syntactic structures seem to be more restrictive with regard to countability (e.g. Fr. *quelques* ‘some_{PL}’ vs. *plusieurs* ‘various’) (cf. Lauwers 2014).

After these refinements, I can now present my integral model of collection nouns. Since I assume that semantic-syntactic characteristics roughly reflect an underlying conceptual structure, I have chosen a graphical representation for my model of collection nouns which is able to reflect all these different aspects. The idea of a graphical representation of nominal types of this kind is not new and can already be found in Mihatsch (2000) or Joosten (2006; 2010); my own model as depicted in Figure 3.1 below consequently follows their methodological approach. I also follow Mihatsch (2000, 39) and Joosten (2006; 2010) in assuming collection nouns as being part of an aspectual continuum leading from the reference to singular objects to collections to plural objects. This continuum is represented in the following Figure 3.1 and will shortly be explained.

As shown in Figure 3.1, the two poles of singular and plural object nouns represent an aspectual dichotomy. Singular object nouns denote bounded entities, which may or not be made of various parts, but these constituting parts or entities are conceptually and linguistically irrelevant and inaccessible (cf. Lakoff 1977, 246–247; Jackendoff 1991a, 20–21; Croft/Cruise 2004, 63–64). On the other end of this aspectual continuum, plural object nouns denote an unbounded plurality of bounded entities, with no cohesion between them (cf. Jackendoff 1991b, 20–21; Meisterfeld 1998, 45). I adopt this terminology of Rijkhoff (2002) to refer to bounded, non-homogeneous referents to avoid notions like *singular/plural count noun*, which may well also refer to count nouns denoting collections. Other terminological choices like *individual noun* may also be possible. In between these two aspectual poles are the different kinds of collection noun whose characteristics were elaborated in chaps. 3.1.1 and 3.1.2. As indicated, it is not only in the context of these descriptions that there are significant differences between animate and

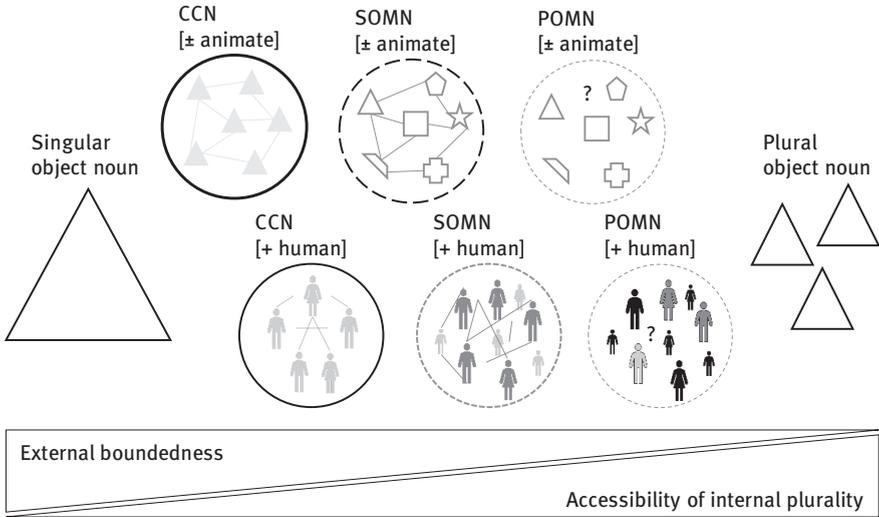


Figure 3.1: Schematic representation of different types of collection noun.

inanimate referents on the one hand and human referents on the other. This distinction is marked by the two different types of each collection so that there are two series of collection nouns representing them displayed in Figure 3.1. Beginning with non-human CCNs, the previous discussion of the state of the art has clearly shown that these designate bounded collections which comprise homogeneous, hardly accessible entities. A typical case for this aspectual type is e.g. Fr. *bouquet* ‘bunch of flowers’. In contrast, human CCNs show a slightly lesser degree of external boundedness, marked with the thinner outline, and therefore an equally slightly higher degree of accessibility of the constituting entities by means of e.g. constructions *ad sensum*. This linguistic means of accessing internal plurality is, however, language-specific. As shown in chap. 3.1.2, human CCNs are not only structured by meronymy, but also mostly characterised by a structured hierarchy with different levels of organisation. This is the case of Fr. *équipe* ‘team’ or also Fr. *jury* ‘jury’. The next category is represented by non-human SOMNs like Sp. *ropa* ‘clothing’, Fr. *vaisselle* ‘crockery/dishes’ or Engl. *furniture*. In contrast to non-human CCNs, non-human SOMNs denote highly heterogeneous entities, which are linked by both meronymic and hyponymic relations. The degree of external boundedness again is smaller corresponding to a higher degree of accessibility of the constituting entities. Note that this linguistic access is only possible with the aid of lexical means of expression, such as stubbornly distributive predicates, but not through grammatical means, such as the direct combination with any kind of distributive determiner. With human referents,

the constituting entities are additionally accessible through syntactic means of expression like *constructions ad sensum* or plural anaphoric pronouns. The limited amount of research on POMNs only allows for speculations, but the aspects discussed so far let to assume that non-human POMNs like *groceries* or Fr. *vivres* ‘food supplies’ and human POMNs like Fr. *gens* or Germ. *Leute* ‘people’ are to be classified under other types of collection noun in accordance with the tendencies displayed, viz. increasing degree of accessibility of the internal plurality and decreasing degree of external boundedness. There is uncertainty regarding the extent to which the constituting entities of a POMN are linked, whether they are indeed linked at all and, if this is the case, what kind of relation they are subject to. Although the schematic representation implies a clear delimitation of nominal categories, these are assumed rather to represent a continuum with similarities and differences and representatives of each category tending more to the prototype or to the periphery of the category (cf. Gardelle 2017 for a similar line of argument). Related to this is also the question, in how far homogeneous mass nouns as well as nouns denoting granular aggregates would fit into this schematic representation. As indicated in chap. 2, these two aspectual types are not in the centre of interest of this present work, but the theoretical framework adopted here indicates that they might be integrated easily. Following my line of argument, mass nouns like *water* construe their referents as maximally unbounded as well as maximally homogenous. Furthermore, as elaborated above, granular aggregate nouns may parallel with OMNs in their aspectual characteristics, but I assume that their degree of accessibility of the internal plurality is generally lower, due to the low degree of saliency of the constituting entities (e.g. grains of rice). With respect to their degree of external boundedness, they might thus be located somewhere between the two poles of homogeneous mass nouns and plural object nouns. Moreover, Mihatsch/Kleineberg (in prep. a) show that the degree of accessibility indeed depends on the fossilised morphological number, i.e. singular or plural. It follows from this that a two-dimensional schema would be more appropriate if one wants to include these aspectual types: we may then take the degree of boundedness as well as the degree of accessibility of internal plurality as the x- and y-axes. Homogeneous mass nouns would then represent point zero where both axes intersect, the other nominal types would then be distributed in accordance to their specific combination of both x- and y-values. An approach similar to this theoretical sketch is also adopted by Ising (2019, 81).

This initial approach of an integral semantic-syntactic model of collection nouns still leaves open a number of questions that need to be addressed in empirical research. These correspond to the main research question (RQ1) formulated in the introduction: What are the influencing factors on the particular linguistic

expression of a collection of entities and the semantic-syntactic characteristics related to it?

RQ_{ss}1: Given the assumed continuum of aspectual types as assumed in Figure 3.1, to what extent do morpho-syntactic features like countability and morphological number correlate with conceptual-semantic aspects like the constitution of the internal plurality of a collection or the possible linguistic accessibility of it?

RQ_{ss}2: CCNs are well studied, but Romance OMNs have up to now not been in the focus of interest in Romance linguistics. To what extent do the characteristics delimited for the English language, summarised in chap. 3.1, also hold for Romance? To what extent does the language-specific nominal system, summarised in Table 1.2, influence the semantic-syntactic properties of Romance OMNs?

RQ_{ss}3: Given the assumed continuum of saliency of constituting referents, i.e. human vs. non-human referents, to what extent does the ontological type of referents correlate with the linguistic accessibility of them?

Combining the assumptions of Figure 3.1 and the summary of the different nominal systems in Table 1.2, I assume the three features of language-specific nominal system, collection noun type and ontological type of referents to correlate. These research questions regarding the semantic-syntactic characteristics of collections will be addressed in the empirical study presented in chap. 4 in order to base this integral model of collection nouns in Romance on experimental grounds. These analyses should also allow for a refinement of the model, which will be presented in a revised version in chap. 4.2.3.

3.2 Collection nouns as a morphological category

As has been shown, the majority of research on collection nouns in general adopts a semantic-syntactic perspective, and the morphological form mostly only becomes relevant when countability comes into play (cf. chap. 3.1.1). As a consequence, derivational processes of collection nouns have been barely investigated, and when these are the focus, it is mostly only when French is the language of analysis (cf. i.a. Aliquot-Suengas 1996; Dubois/Dubois Charlier 1996; Flaux 1999; Gross 2011; Mihatsch 2021). This neglect of the morphological form of collection nouns is due to the apparent absence of a suffix in Romance languages that exclusively forms this nominal type. Michaux (1992, 105), who tries to find common characteristics of collective nouns, summarises this simply: “[. . .], the morphological tests do not help us [to delineate this category]” (cf. also Borillo

1997, 114). Considering also the derivational processes involved in the creation of collection nouns, most studies are thus restricted to mere lists of collective suffixes (cf. Aliquot-Suengas 1996, 36–37; Solé Solé 2002, 63) and even in these lists, the suffixes are not classified uniformly, as has been well illustrated by Solé Solé (2002, 69). In what follows, I will show that considering collection nouns as a Romance word-formation technique is nevertheless indispensable: although most Romance collective suffixes are indeed highly polyfunctional, this polyfunctionality is not arbitrary but systematic. The analysis of the derivational processes involved in the forming of collection nouns has to be consequently regarded as a – if not necessarily obligatory, but still – optional criterion for the definitional delimitation and linguistic description of collection nouns.

I have already described the different types of collection noun that are governed by their semantic-syntactic characteristic and, taking into consideration their morphological form, we can also differentiate between various categories. First of all, one has to distinguish between morphologically simple (e.g. Fr. *famille* ‘family’, or Sp. *gente* ‘people’) and complex collection nouns mostly derived through suffixation (e.g. It. *vasellame* ‘crockery/dishes’, Sp. *amueblamiento* ‘furnishings’). The derivational process of compounding is not a common way in Romance languages to form collection nouns, but it is very widespread in Germanic languages – the latter will be addressed after the consideration of Romance for a small-scale cross-linguistic comparison. Possible compound collection nouns are e.g. Germ. *Schuhwerk* ‘footwear’ or Engl. *kitchenware*. Complex collection nouns in Romance can be additionally differentiated into two further categories: on the one hand, there are collection nouns which have only become such through processes of semantic change (e.g. Sp. *armamento* ‘arming/weaponry’, *-mento* being a predominantly eventive suffix, cf. DES, s.v. *-miento*, *-mento*, *-menta*) while, on the other hand, there are collection nouns derived with the aid of collective suffixes (e.g. Fr. *valetaille* ‘servants’). As illustrated by the Spanish example of *armamento*, many collection nouns derive from deverbal action nouns that receive the additional collective meaning through metonymic shifts. In contrast, collection nouns directly derived with the aid of collective suffixes are the result of a clustering of referents expressed by the base noun (*valet* ‘servant’ > *valetaille* ‘group/totality of servants’). In the following, I will discuss these three types in more detail, concentrating on the state of the art on Romance languages and on English and German as reference points for cross-linguistic comparison. Chap. 5 will further address the question whether the deriving of collection nouns can count as a productive word-formation process in Romance languages, focusing mainly on French, Spanish and Italian. Before coming to the analysis of the morphological processes involved in the forming of collection nouns, a short note on typology is necessary. As will be shown in what follows, collection

nouns in Romance languages, as well as in English and German, are in most cases the marked form, but this is not necessarily the case in all languages. Arabic, for example, has a category of generic collection nouns that has a morphologically marked singulative equivalent: *samak* ‘fishes/fish (as a species)’ vs. *samak-a* ‘(a single) fish’, and Welsh distinguishes between an unmarked plural and a marked singulative form like *dail* ‘leaves’ vs. *dail-en* ‘leaf’ (cf. Greenberg 1977, 288; Kuhn 1982, 61; Acquaviva 2015; Haspelmath/Karjus 2017). Maltese furthermore distinguishes between the two formal categories of *singular* and *non-singular*, the latter being the unmarked one: *Pietru għandu kartun* ‘Peter has cardboard’ (non-singular/mass) vs. *Pietru għandu kartun-a* ‘Peter has a cardboard’ (singular) (cf. Gil 1996, 71–72). It follows from these examples that the description of Romance and some Germanic languages cannot be taken as the default but has to be viewed as characteristic for this special linguistic type.

Starting with the consideration of morphologically simple collection nouns, even if we synchronically consider these as not being products of word-formation processes, many of these can be traced back to phenomena such as fossilisation of inflectional morphemes. This is the case with the Romance continuations of Lat. *FAMILIA* ‘family’ which originally designated the totality of *famuli* ‘servants’, *-ilia* is therefore the fossilised inflectional suffix that marks the neuter plural (cf. Gaide 1989, 225). Another example is Fr. *vaisselle* ‘dishes’ which goes back to the Latin neuter plural *VASCELLA* ‘small containers’ (cf. Mihatsch 2006, 131). There is also some kind of derivational process *lato senso* in the evolution of Sp. *ropa* ‘clothing’, which goes back to the Germanic verb *RAUBÔN* ‘to rob’. In its origins, *ropa* designated the result of this action, i.e. ‘what has been robbed’, and was first established as ‘booty’ and then ‘clothing’ as it was an especially valued type of loot (cf. DECH, s.v. *robar*, cf. also Mihatsch 2016, 292). Furthermore, there are cases of lexical absorption, like Fr. *biens mobiliers* ‘movable property’, where the adjective has been converted into the noun *mobilier* ‘furniture’ (cf. TLFi, s.v. *mobilier*; cf. also Engl. *mail* < *mail of letters*, cf. OED, s.v. *mail*). Generally, these kinds of word-formation process are less relevant to Romances languages or to English or German. More importantly, Romance collection nouns are mostly derived through suffixation (cf. Mihatsch 2021), whereas English and German make use of a series of collectivising suffixoids. I will first concentrate on Romance and then on English and German to determine more general tendencies.

As indicated in the introductory remarks, there are two types of Romance suffix that represent the derivational basis for collection nouns.²³ On the one

²³ Grossmann (2004) gives an overview over Italian collective suffixes; Bosque (1999, 33) concentrates on Spanish, the *Nueva Gramática de la lengua española* treats Spanish collective suffixes in the framework of a general description of the Spanish suffix inventory (cf. Real Academia Española/Asociación de Academias de la Lengua Española 2009, chap. 6; cf. also Rainer 1993,

hand, there are suffixes that are highly polyfunctional with the collective function therefore being mostly the result of a metonymic shift, on the other, there are suffixes that genuinely form collection nouns. Both kinds of suffix have a Latin origin and neither of the two types already had a collectivising function in its origins. Collectivising was a rather marginal derivational category in Latin, because the plural was much more flexible than in Romance languages: “Virtually any Latin noun can be pluralized, and in fact frequent occurrences of plurals of ‘mass-denoting nouns’ or abstract nouns [. . .] are attested” (Stark 2007, 56). In this respect, special attention has to be paid to the neuter plurals formed with the suffix *-a*. On the one hand, it is the regular neuter plural form (*folium – folia* ‘leave – leaves’)²⁴ but, on the other, it can also function as a collective plural, as opposed to the regular distributive plural: *locus* ‘place, location’ – *loci* ‘places, locations’ – *loca* ‘area (= collection of places)’ (Schön 1971). “Es [das Suffix *-a*] ist nicht so sehr neutrales Pluralmorphem, als vielmehr *genusindifferentes Kollektivmorphem*, [. . .]” (‘it [the suffix *-a*] is not a suffix marking the neuter plural, but much more a gender-neutral collective morpheme’) (Schön 1971, 56).

Derivational collective suffixes, i.e. those not related to the inflectional neuter plural suffix, were mostly limited to the semantic domain of plants and trees like Lat. *olivētum* ‘olive tree plantation’ or Lat. *arborētum* ‘grove’. These were either not continued, are still restricted to the domain of plants, or they had already lost their productivity in older stages of development of Romance languages. Lat. *-ĒTUM* e.g. still survives in Sp. *robledo* ‘oak grove’ or Pt. *ulmedo* ‘elm grove’, the French suffix *-oi* however only survived until Old French and then disappeared (e.g. OFr. *sapoi* ‘pine forest’ > MFr. *forêt de sapins* ‘idem’) (cf. Gaide 1989; Lüdtke 2007, 323–324). Turning back to Romance languages, it can be consequently stated that the Romance collective suffixes have no collective origin in Latin. According to the distinction introduced above and following Baldinger (1950), the *metonymic* suffixes mostly originate in Latin eventive suffixes, whereas those suffixes which directly cluster the underlying referents like Fr. *-aille* go back to the Latin neuter plural. Table 3.1 gives an overview of Romance suffixes which have a primary or secondary collective meaning. It focuses on the most frequent collective suffixes, leaving out e.g. Sp. *-al/-ar*, because these are especially limited to the semantic domain of locations of certain plants or animals (e.g. *palomar* ‘dovecote’, *arrozal*

206–207); Baldinger (1950) as well as Aliquot-Suengas (1996) deal extensively with the French language. See Mihatsch 2021 for an up-to-date Pan-Romance comparison.

²⁴ In many cases, the feature of plurality however disappears as in Fr. *feuille* ‘leaf’ (fem. sg.) or Sp. *hoja* ‘leaf’ (fem. sg.) from Lat. *FOLIA* ‘leaves’ (neut. pl.) – and had to be re-established through suffixation (cf. Fr. *feuillage* or It. *fogliame* ‘foliage’; cf. Schön 1971, 94–99). See Ising 2019, 301 for the paths of reconceptualisation (in his terminology) involved in this process.

‘rice field’; cf. DRAE, s.v. *-al*; s.v. *-ar*), which in addition should be defined as collection nouns only to a limited extent. I also consider, in addition to French – the focus of the present work – Spanish and Italian. Spanish may be interesting, because it is said to have the collective suffix *-erio* which is mostly productive in the Mexican variety, but which until now has not been considered in studies on derivational patterns in Spanish (cf. Ponce de León 2016). Italian is included, since teenager’s usage patterns in particular seem to show productive collectivising processes (cf. Poletto/Penello 2006; Magni 2018). Since the Portuguese and Catalan suffix inventories potentially indicating collection nouns are similar to the Spanish one, they will not be of further focus in the following overview. Characteristics diverging from the Ibero-Romance neighbours will be commented when necessary (cf. Correia 1999; Rio-Torto/Soares Rodrigues 2016, 153–157 for an overview of collective suffixes in Portuguese; cf. Solé Solé 2002, chap. 3 for a critical overview of collective suffixes in Catalan). For every suffix, the main meaning nuances are indicated, following their degree of prototypicality (according to lexicographic information) and reducing them to principal categories like *action* or *collection*. The overview is completed by an exemplifying collection noun based on the respective suffix.²⁵

Table 3.1 illustrates three possible paths of development of Romance collective suffixes. First of all, there are the Latin eventive suffixes which in general lead to highly polysemous Romance suffixes, and where the collective function here is only secondary. Secondly, there are those suffixes that go back to the Latin neuter plural, where here the collective function can be primary (cf. *-ĀLIA*). Finally, there are two Latin suffixes marking relational adjectives (*-ĀTICUM/-ĀRIUS*) which represent the basis of a rather diverse suffix category in Romance, but which also have collectivising potential. This potential lies in the fact that relational adjectives like Lat. *viāticum* ‘related to travel’ may imply a plurality that in turn may lead to the formation of a collection noun. In what follows, I consider these three categories in more detail to thereby return to the original question of whether the morphological form of a collection noun can serve as a definitional criterion for this category.

²⁵ The overview is based on Baldinger (1950), Grossmann (2004) as well as Real Academia Española/Asociación de Academias de la Lengua Española (2009) and was supplemented by lexicographic information (Fr.: TLFi; Sp.: MM, DRAE; It.: Zingarelli). Individual references: Lat. *-MEN* and its continuations in Romance languages: Rainer (2018); *-ĀRIUS/-A* and its continuations in Romance languages: Rainer (2018).

Table 3.1: Romance collective suffixes and their semantic nuances.

Latin	French	Spanish	Italian
Action – effect	- <i>ain</i> / - <i>in</i> / - <i>un</i> : <i>fretin</i> 'swarm of small fish' Collective function restricted, not productive	- <i>ambre</i> / - <i>amen</i> : <i>maderamen</i> 'woodwork' Collection	- <i>ame</i> / - <i>ume</i> : <i>fogliame</i> 'foliage' Collection (mostly pej.)
	- <i>(t)ure</i> : <i>scriptura</i> 'writing/document' Action, effect	- <i>dura</i> : <i>dentadura</i> 'denture' Action, effect, instrument, collection	- <i>tura</i> : <i>attrezzatura</i> 'equipment' Action, effect, collection
	- <i>MENTUM</i> / - <i>MENTA</i> : <i>alimentum</i> 'food' Means, instrument	- <i>menta</i> / - <i>mienta</i> : <i>osamenta</i> 'skeleton' / <i>herramienta</i> 'tool/toolkit' Collection	- <i>mento</i> : <i>abbigliamento</i> 'clothing' Action, effect, collection
Neuter plural	- <i>at</i> < - <i>ĀTUM</i> : <i>professorat</i> 'teachers' State, function, office, collection - <i>ée</i> < - <i>ĀTA</i> : <i>cochonné</i> 'litter of piglets' Action, collection, unit of measure - <i>ade</i> < - <i>ĀTA</i> : <i>colonnade</i> 'colonnade' Action, collection	- <i>ada</i> / - <i>ada</i> : <i>campesinado</i> 'peasantry' / <i>vacada</i> 'herd of cattle' (deverb.) Action, effect (denom.) Collection, office (period), place of office	- <i>ata</i> : <i>fiaccolata</i> 'torchlight procession' (deverb.) Action, effect (denom.) Collection, unit of measure, time period - <i>ato</i> : <i>tavolato</i> 'board wall' Collective function restricted

Table 3.1 (continued)

Latin	French	Spanish	Italian
-ĀLIA: <i>animālia</i> 'animals' Neuter pl. of -ālis	-aille: <i>pierraille</i> 'gravel' Collection, pejorative	-alla: <i>morralia</i> 'junk' Collection, pejorative	-aglia: <i>ferraglia</i> 'scrap metal' Collection (mostly pej.)
Relational -ĀNTICUM: adjectives <i>viāticum</i> 'related to a voyage'	-age: <i>feuillage</i> 'foliage' (deverb.) Action (denom.) Collection	-aje: <i>plumaje</i> 'plumage' (deverb.) Action, effect (denom.) Collection	-aggio: <i>baronaggio</i> 'barony' Action, effect, collection
Relational adjective	-ier + -ie > -erie: <i>lingerie</i> 'underwear' Characteristic, action, place, collection	-ario/-aria: <i>poemario</i> 'collection of poems' Profession, place, beneficiary, collection	-eria: <i>armeria</i> 'weaponry' Profession, collection, characteristic
-ĀRIUS/-IA: <i>sumptuārius</i> 'sumptuary' Relational adjective		-ero + -ia > -eria: <i>palabrería</i> 'idle talk' Collection, characteristic, action, place	
		-ío: <i>gentío</i> 'crowd' Collection, state	
		-erío: <i>caserío</i> 'hamlet/farm' Collection	

The Latin suffixes -MEN, -TŪRA and -MENTUM are all characterised by having a more or less clear relation to an action.²⁶ As means of deriving deverbal nouns, they generally designate results, instruments or the action itself. In this respect, Lat. *agmen* (< *agere* ‘to move’) designates the ones who move and Lat. *alimentum* refers to a means to fulfil an action, namely feeding someone or something. The action reference is most clear in the example of *scriptūra* which either designates the action of writing or its result. Nearly all these suffixes show a continuous evolution in the three Romance languages studied here, and only in French is the continuation of Lat. -MEN no longer productive (cf. Baldinger 1950, 143–145; Rainer 2018b, 436–437). Especially in the case of -MEN, the different variants determined through different thematic vowels diverge in Spanish and Italian. In Spanish, it is mainly the variant of *-amen* that is continued in the Latinised form, while the popular form of *-ambre* is no longer productive (cf. Rainer 2018b, 448). In Italian both *-ame* and *-ume* are continued as possibly collectivising suffixes (mostly in combination with nominal and adjectival bases respectively), and *-ime* is very rare in Standard Italian (cf. Rainer 2018b, 417–420). In addition, Portuguese shows a high variation in its continuations of Lat. -MEN, not only regarding thematic vowels, but also Latinate variants and regional variations like *-um* (cf. Rainer 2018b, 450). In each of these cases, the collective function is derived via a metonymic shift from the resultative or instrumental interpretation of the *nomen actionis* in question. This was already possible in Late Latin as shown by Rainer (2018b, 404–410) for derivatives on -MEN: e.g. Lat. *farcimen* ‘sausage’ stems from the verb *farcire* ‘to stuff’. At this stage of evolution, the possible collective interpretation of a derivative is occasional to non-existent and not inherent to the suffix: “Einen kollektiven Sinn drückt [-]mentum nie aus” (‘-mentum never expresses a collective meaning’) (Roedinger 1904, 16; cf. also Baldinger 1950, 109–110). In Romance languages, this secondary collective function of deverbal nouns is transferred to a primary collective meaning of denominal derivatives. Consequently, the collective interpretation of the suffix evolves “durch den Übergang von der spontanen zur funktionellen Kollektivität, d.h. formal durch den Übergang von der deverbale zur denominalen Ableitung” (Baldinger 1950, 215) (‘through the transition from the spontaneous to the functional collectivity, i.e. formally through the transition from a deverbal to a denominal derivation’). In some cases, the eventive interpretation of the suffix becomes completely lost (cf. It. *-amen*), or an inherent collectivising function of the respective suffix was never fully devel-

²⁶ As noted by Roedinger (1904, 17f.; 59), this eventive function is – at least in the case of -MENTUM – not necessarily the etymologically primary function. This suffix originates as an indicator of the abstract means of an action and only in post-classical Latin increases the number of derivatives relating to the action itself.

oped; this is the case with Fr. *-ment*, where the collection nouns derived through this suffix only got their collective meaning through change of meaning of the derivate (cf. Roedinger 1904, 60).²⁷

The second category of collective suffixes goes back to the Latin neuter plural, the suffixes *-ĀTUM/-ĀTA* representing a transition category to the first, since they too show an action-effect polysemy. In this respect, neither *-ĀTA* nor *-ĀLIA* are feminine singular but neuter plural, and the reanalysis of gender only takes place in transfer from Latin to Romance (cf. DES, s.v. *-ada*; Baldinger 1950, 123). At the beginning of the evolution of Lat. *-ĀTA* and its Romance continuations, there are Classical Latin derivations of the fourth declension like *piscātus*, *-ūs* ‘the action of fishing’ which are reanalysed as derivatives of the second declension in spoken Late Latin: *terraemōtus*, *-ūs* (masc.) > *terraemotum* (neut.) ‘earthquake’. This new form with neuter gender naturally produces equivalent plural forms (*cōgitātum – cōgitāta*, cf. Table 3.1), a general development of abstract and collection nouns in spoken Latin (cf. Collin 1918, 35–40; Stark 2007, 56; DES, s.v. *-ada*). These neuter plural nouns designating actions and events are further reanalysed as feminine singular and thence follow a similar path as the already described continuations of Latin *-MEN*, *-TŪRA* and *-MENTUM*. Through a metonymic shift, the meaning changes from denoting an event either to denoting the people or things involved in this particular event or to denoting more than one instance of it. As a consequence, the French suffix *-ée* still derives collective events like *fessée* ‘spanking’, but also denominal collection nouns like *cochonnée* ‘litter of pigs’ or *tablée* ‘group of guests sitting at the same table’ (cf. Baldinger 1950, 107–109; 207 ff.). Fr. *-ade* is a medieval loan from Italian, Spanish and Occitan (cf. TLFi, s.v. *-ade*). For all three languages studied here, the continuations of Latin *-ĀTA* seem however to derive nouns especially from this first type of collective event, like Sp. *cuchillada* ‘(multiple) stabs’ or nouns denoting the content of something as in It. *cucchiaiata* ‘a spoonful of’. The Romance continuations based on the neuter *-ĀTUM* give a similar picture in being only secondary collectives, but these took a different evolutionary path. Latin *-ĀTUM* mainly derives action nouns that belong formally to the fourth declension as described above. In Romance languages, its continuations of Sp. *-ado* and It. *-ato* however create a new process based on adjectives that imply some plurality.

27 There are marginal cases of derivatives on *-ment* in Old French which have no relation to an action in being denominal and in denoting only a referential collection. Typical cases are for instance *étoilement* ‘collection of stars’ (< *étoile* ‘star’), *vaissellement* ‘crockery’ (< OFr. *vaisselle* ‘recipient’) or *outillement* ‘collection of tools’ (< *outil* ‘tool’) (cf. Roedinger 1904, 111–112). These derivatives thus prove that the suffix *-ment* began to develop the capability of forming denominal, functional collections, but this capability ceased to exist. This impression is also confirmed by my own empirical analysis on nonce-formations (cf. chap. 5.3.1).

In this respect Sp. *almenado* ‘collection of merlons’ e.g. goes back to the adjective *almenado* ‘with merlons’. In this case, the implication of a plurality of referents was already inherent to the adjective and was then transferred to the noun derived from it (cf. DES, s.v. *-ado*). Collective derivates on Fr. *-at* seem to have taken yet another path leading from nouns denoting some function or dignity like Fr. *magistrat* ‘magistrate’ (< Lat. MAGISTRĀTUS ‘idem’) to the people exercising this function (cf. TLFi, s.v. *-at*). Summarising the aspects of Lat. *-ĀTUM*, *-ĀTA* and their continuations in Romance, they – like *-TŪRA*, *-MEN* and *-MENTUM* described above – produce collection nouns only as secondary derivational products. Here again, the change in meaning too is based on a metonymic shift from e.g. the action itself to people or things related to it or a function to the people exercising it.

The suffix *-ĀTA* represents a bridging case between these cases of systematic polysemy and the derivation of collection nouns based on the Latin neuter plural, represented by the continuations of Lat. *-ĀLIA*. Those seem to be the only suffixes in Table 3.1 that evade the systematicity of the polysemy described above. Aliquot-Suengas (1996, 291) comments on Fr. *-aille* (cf. also Baldinger 1950, 83): “Il existe donc en français un suffixe et un seul qui construisent du collectif et le collectif est bien une catégorie dérivationnelle du français, même si une seule marque peut lui être associée” (‘There is thus in French a suffix and only one suffix that derives collective nouns and COLLECTIVITY is indeed a derivational category of French, even if only one mark can be associated with it’).

There is at present a number of collective derivates on Fr. *-aille*, Sp. *-alla* or It. *-aglia*, often but not necessarily with an additional pejorative connotation: Fr. *valetaille* ‘servants’, *poulailler* ‘poultry’, *feraille* ‘scrap metal’, Sp. *canalla* ‘rabble’, *morralia* ‘junk’, It. *ferraglia* ‘scrap metal’, *ragazzaglia* ‘bunch of urchins’ etc. Characteristic of these derivates is their feature of mostly being mass nouns: “Avec les dérivés en *-ail(le)*, on a donc affaire à des *ensembles massifs* dont les parties, atomiques ou complexes, en morceaux ou non, sont soumises à une opération de désindividualisation” (‘With the derivates on *-aille* we thus have *mass collectives* of which the parts, atomic or complex, in chunks or not, are subjects to an operation of deindividualization’) (Aliquot-Suengas 1996, 139; my stress).

At this point, a more detailed note on the relation between COLLECTIVITY and pejorative connotation is necessary. As already stated above, many collection nouns on Fr. *-aille* e.g. are said to overwhelmingly have a negative connotation, and this is said to be mostly the case with human derivates (cf. Baldinger 1950, 126–127). The relation between a pejorative character and COLLECTIVITY is not limited to this special suffix, but can be found in many collective derivates, especially when the constituting referents are human (cf. Mihatsch 2015a, 1190). For Italian, this strong relation is assumed especially for the suffixes *-ame*, *-ume* and *-aglia* (e.g. *gentame* ‘rabble’, *tedescume* ‘bunch of Germans’, *ragazzaglia*

'bunch of urchins'; cf. Grossmann 2004, 245–247), while for Spanish and French nearly all collective suffixes are said to have the potential to create a negative connotation (cf. Baldinger 1950, 221–222; Rainer 1993, 207). The especially strong tendency of Fr. *-aille*, but also It. *-aglia* and Sp. *-alla* to derive pejorative collection nouns could be due to their mass character: through the massification of the single constituting elements, the aspect of individuality is even more reduced (cf. Mihatsch 2015a, 1190). Another possible explanation, especially for collection nouns going back to Lat. *-MEN*, is the fact that these suffixes first of all derived collection nouns which comprise animals or things as referents. Human collection nouns based on these suffixes only appear later in history and are therefore metaphorical extensions of nouns like Lat. *pullāmen* 'young animal(s)'. Consequently they innately have the potential for pejorative connotations (cf. Rainer 2018b, 454). This also holds for the collection nouns that go back to Lat. *-ĀLIA* which probably are mostly derived analogously to Lat. *animālia* 'animal (adj. neut. pl.)' (cf. Baldinger 1950, 130–133).

The last group of derivational collective suffixes comprises the Romance continuations of Lat. *-ĀTICUM* and *-ĀRIUS* which derive relational adjectives and deadjectival nouns based on those. The first suffix primarily marked relational adjectives like *viāticum* 'relative to travel' or *domesticum* 'domestic/relative to the household'. The first example probably functioned as a focus of attraction for the evolution of a derivational morpheme, just as in the case of the continuations of Lat. *-ĀTUM* (cf. *supra*) in Gallo-Romance. The collective suffixes in the other Romance languages are not direct continuations of the Latin origin but represent loans from Gallo-Romance (cf. Lüdtke 2007, 325; DES, s.v. *-aje*). It can no longer be reconstructed whether the plural meaning associated with these kinds of adjective is transferred to the respective noun via nominalisation or by ellipsis of the adjective – both options may be possible.

The evolution of the second group of suffixes going back to Lat. *-ĀRIUS/-A* is slightly more complex. In Latin, it derived relational adjectives like *sumptuārius* 'related to expense' and represents the basis for Romance suffixes like Fr. *-aire* and *-ier* as well as Sp. *-ero* and It. *-iere*. These suffixes derive relational adjectives (e.g. Fr. *financier* 'financial'), but also agent nouns (e.g. It. *infermiere* 'nurse') and locatives (e.g. Sp. *hormiguero* 'anthill'), and nouns denoting collections are harder to find with this type of suffix or are regarded as rare and/or obsolete (e.g. It. *ladronaia* 'band of robbers'; cf. Rainer 2018a). Through combination with the respective Romance continuations of Lat. *-IA*, however, a series of highly poly-functional suffixes have evolved that equally derive nouns denoting collections, offices and professions, locations as well as states and actions. An illustrative example is Fr. *bijouterie* which denotes the action of making jewellery (action), the shop where it is sold (location) as well as the jewellery itself (collection) (cf.

Lüdtke 2007, 327; TLFi, s.v. *bijouterie*). Consequently, these types of derivation also follow the systematic polysemy already described between an abstract event and the concrete artefact related to it (cf. Baldinger 1950, 175; 206 for a detailed overview of the abstract and concrete functions of Fr. *-age* and *-erie*). In Spanish, we see another development: the Latin suffix *-IVUM*, also deriving adjectives like *laudātīvus* ‘laudatory’, has evolved in Spanish to the suffix *-ío* which derives collection nouns (e.g. *gentío* ‘crowd’), as well as states (e.g. *monjío* ‘state of being a nun’) and authorities (e.g. *señorío* ‘leadership’). Probably analogously to *-ería*, *-ío* also evolves the variant *-erío* which only has a collectivising function (e.g. *caserío* ‘farmstead’) (cf. Pharies 2004, 163; Ponce de León 2016, 32–34). Because of these analogies and parallels, I classify Sp. *-ío* and *-erío* together with the direct descendants of Lat. *-ĀRIUS*.

After this description of the single Latin and Romance suffixes, some general tendencies can be summarised. First of all, “only for a few of them is the formation of collectives the primary function” (Magni 2018, 212). Repeating what was said at the beginning of this chapter, two groups of derived collection noun can be determined: on the one hand, those lexemes that were derived with a non-collective suffix and got their collective meaning only through a change in meaning and on the other hand, those that were derived with a collective suffix. I therefore conclude, basing my assumption on Baldinger (1950), that deverbal nouns are in general are not collection nouns in the first place but only become so through semantic development, and that denominal nouns generally are more likely to have a primarily collective meaning. This generalisation is based on the assumption of spontaneous vs. functional collectivity (cf. *supra*). This also applies to suffixes that have their roots in relational adjectives and that originally only had the possibility of a non-collective meaning, but which also gained their functional collectivity at the point when they could be productively combined with nouns, directly collectivising the underlying referents. The zero-derived nouns based on these suffixes like Lat. *viāticum* ‘related to travel (adjective)/provisions for a journey (noun)’ may have supported these changes in meaning. In those cases where the collective meaning of the suffix only arises by (metonymic) meaning change, one might assume that this change first occurred in the case of individual derivatives like Lat. *viāticum*. The more often this meaning change is associated with a certain suffix, the more entrenched becomes this meaning in the suffix itself. Formally, this shift is manifested by the change from deverbal and deadjectival derivatives to denominal ones (cf. Mihatsch 2021).

Consequently, summarising the state of the art, there seems to be broad agreement on the fact that collective suffixes are highly polysemous. Even Baldinger (1950, 7), who dedicated a whole monograph on research into collective suffixes in French, has to admit: “Der Begriff ‘Kollektivsuffix’ ist an sich schon problema-

tisch, denn es gibt kaum Suffixe, die in allen Verhältnissen, unter allen Umständen nur kollektive Ableitungen bilden” (“The term *collective suffix* is problematic in itself because there are hardly any suffixes that only form collective derivatives in all circumstances”). Prototypical examples of this type are Fr. *-(t)ure*, Sp. *-aje* and It. *-ata*. Furthermore, Romance continuations of Lat. *-ĀLIA* all seem to represent one suffix which exclusively derives collection nouns. These are characterised by often having pejorative connotations and by being mostly OMNs. Additionally, these are said to be no longer productive. Consequently, this summary of the state of the art leads to the conclusion that collection nouns do not appear to represent a very productive category of word-formation in Romance languages. In chap. 5, I will address whether this really is the case when I adopt the theoretical perspective of construction morphology and analyse collective nonce-formations in French, Spanish and Italian.

I conclude this chapter by making a detour to the Germanic languages, represented here by German and English. These languages make use of a series of suffixoids, which are neither full suffixes nor compositional elements. Canonical examples are Germ. *-zeug*, *-schaft*, *-kram*, *-werk* and Engl. *-wear*, *-gear*, *-ship*, *-hood* or *-dom* (cf. Plag 2003, 86–92; Mihatsch 2015a, 1187–1188).²⁸ All these suffixoids originate in fully fledged nouns which had the most diverse meanings, although for many of these their status as an independent noun was dying out or had already been lost in Middle High German (sometimes already in Old High German) and Middle English (cf. Erben 1959, 224; 2003, 2531; Meineke 1991, 117; Meineke 1994, 521; Trips 2009). They share the property that the collectivising function was not their original one, but only developed through a metonymic shift – here we consequently find a clear parallel to Romance languages²⁹ (cf. Rosenkranz 1968; Erben 2003, 2531; Trips 2009). A collection noun like Germ. *Schuhwerk* ‘footwear’ e.g. originally had the meaning of ‘shoemaking/craftmanship’, its eventive character being based on *-werk* which has its origins in the verb *wirken* ‘to work’ – the collective function only evolved later (cf. GDW, s.v. *Schuhwerk*; cf. also Rosenkranz 1968, 225–227). Germanic derivatives consequently also show the pattern of a concrete realisation of an action (cf. also Germ. *Flechtwerk* ‘wickerwork’/*Täfelung* ‘panelling’), resembling their Romance counterparts also in their polyfunctionality. In this respect, Engl. *-hood* e.g. also derives states like *parenthood* or *neighbourhood* and Germ. *-werk* marginally still has its original

²⁸ In addition, German has the special property of a collectivising prefix *Ge-* as in *Gebrüder* ‘brothers’ or *Gebälk* ‘timberwork’ (cf. Erben 1959, 222). For this property being unique in the group of languages analysed here, it will not be treated any further in this present framework.

²⁹ There is however no sign of grammaticalisation of compositional elements in Romance languages (cf. Mihatsch 2015, 1188).

eventive meaning as in *Hexenwerk* ‘witchcraft’. This detour thus indicates that there seems to be some more or less cross-linguistically uniform principles of deriving collection nouns, at least in the languages treated in this context.

Some of the collection nouns derived through these typical suffixoids represent OMNs and not bounded collective nouns. In this respect Zifonun (2012, 122) finds that “die drei aus Kompositionsbestandteilen grammatikalisierten Kollektivmarker des Deutschen *-gut*, *-werk*, *-zeug* grundsätzlich Kollektivsubstantive ohne Gegenstandscharakter, somit Kontinuativa [erzeugen].” (‘the three collective markers of the German *-gut*, *-werk* and *-zeug*, which are grammaticalised compositional heads, fundamentally derive collective nouns without object character and thus mass nouns’). As in the case of the *mass suffixes* Fr. *-aille*, It. *-aglia* and Sp. *-alla*, these German suffixoids also seem to be specialised for a certain type of collection, namely the unbounded kind. They also have in common the tendency to negatively connote their underlying referents, at least in the case of human ones. Erben (1959, 226) mentions in this respect the example of Germ. *Krankengut* ‘totality of patients’ which objectifies its human referents as something to deal with – this word is not in common usage but is applied in a specialised way to administrative contexts.

This chapter has shown that there are various means of deriving of collection nouns in Romance and Germanic languages. The majority of collectivising suffixes share the property of being highly polysemous between an event-result ambiguity, whereby some suffixes show preferences for deriving OMNs and some appear to show a tendency to derive bounded CCNs. The overall opinion on collection nouns as a derivational category is that the ambiguous character of collective suffixes in Romance languages leads to the conclusion that word-formation patterns do not seem able to serve as a definitional criterion for these nominal types. In contrast to this observation, the semantic-syntactic properties of collection nouns as described in chap. 3.1 seem to be the crucial defining aspects. The Italian data, as described by Mauri (2017) and Magni (2018), in particular suggest that a more updated analysis of these derivational processes is worthwhile. Especially the fact that the great majority of research on this topic has been based only on lexicographic information calls for more quantitative empirical methods to be applied. Chap. 5 will provide new empirical data to address the second research question (RQ2) formulated in the introduction: To what extent are there any productive word-formation patterns in the domain of COLLECTIVITY? This will also analyse the degree to which various sub-types of collection noun display different derivational preferences. In other words, do OMNs represent a derivational category separate from CCNs or do their special syntactic-semantic characteristics only come into being by way of diachronic evolution? This question will be addressed by combining the results of chaps. 5 and 8.

II The synchronic characteristics of collection nouns in present-day language

Part I of this study focused on the state of the art with respect to nominal aspectuality in general and on collection nouns with their semantic-syntactic as well as morphological characteristics. The questions and issues raised in this first part will now be empirically addressed, concentrating on the two primary domains of semantic-syntactic and morphological characteristics. In this part, I will focus on my own empirical studies which analyse the synchronic characteristics of collections in present-day language, and part III will then concentrate on empirical studies regarding diachronic development.

4 Semantic-syntactic characteristics: An acceptability judgement study on count collective nouns and object mass nouns

This first empirical part of the present thesis will focus on the semantic-syntactic characteristics of Romance collection nouns. As explained in chap. 3.1, collection nouns are conceptually defined on the basis of two criteria – the external (un) boundedness and the internal plurality of discrete entities. As was elaborated in chap. 3.1.2, the latter criterion can be further refined by delimiting the respective type of lexical hierarchy which relates the collection to its constituting entities. In what follows, I will elaborate on these aspects for the different types of collection noun focusing mainly on typical spatio-temporally contiguous, bounded CCNs and unbounded OMNs in their morphological singular and plural forms (SOMNs and POMNs respectively). Basing the analysis mainly on an acceptability judgement study, I will now address in more detail my first research question, which regards the influencing factors of the linguistic expression of COLLECTIVITY in Romance languages (RQ1). This comprises firstly, a systematic comparison of different sub-types of collection noun and secondly, a systematic comparison of similar collection nouns in various Romance languages. The state of the art as described in the theoretical preliminaries gives rise to three research questions in this respect, which I repeat here for convenience:

RQ_{ss}1: Given the assumed continuum of aspectual types as assumed in Figure 3.1, to what extent do morpho-syntactic features like countability and morphological number correlate with conceptual-semantic aspects like the constitution of the internal plurality of a collection or the possible linguistic accessibility of it?

RQ_{ss}2: CCNs are well studied, but Romance OMNs have up to now not been in the focus of interest in Romance linguistics. To what extent do the characteristics delimited for the English language, summarised in chap. 3.1, also hold for Romance? To what extent does the language-specific nominal system, summarised in Table 1.2, influence the semantic-syntactic properties of Romance OMNs?

RQ_{ss}3: Given the assumed continuum of saliency of constituting referents, i.e. human vs. non-human referents, to what extent does the ontological type of referents correlate with the linguistic accessibility of them?

Assuming that the three features of language-specific nominal system, collection noun type and ontological type of referents do correlate, as well as in light of the

state of the art, three hypotheses regarding these research questions may now be formulated:

HP_{ss}1: As assumed in Figure 3.1, I expect the acceptability judgement study to show a correlation between an increasing boundedness and a decreasing accessibility of the constituting entities of a collection, and vice versa. In this respect, I assume e.g. CCNs not to be compatible with stubbornly distributive predicates like *round*.

HP_{ss}2: Since English as well as Romance languages are number marking languages clearly distinguishing between a singular meaning ‘one’ and a plural meaning ‘more than one’, I expect Romance OMNs to basically behave similarly to what has been stated for English (e.g. with respect to incompatibility with the indefinite article). But, given the differences in marking the mass-count distinction of the various Romance languages, as summarised in Table 1.2, I expect Brazilian Portuguese OMNs to behave more flexibly than French OMNs with regard to syntactic mass noun properties; European Portuguese, Spanish and Italian may lie between these two poles.

HP_{ss}3: As assumed in Figure 3.1, I expect the acceptability judgement study to show a correlation between an increasing degree of animacy and an increasing degree of accessibility of the constituting referents of a collection noun. In this respect, I assume e.g. inanimate OMNs not to be compatible with highly distributive predicates, but human OMNs may indeed be compatible with such constructions.

In what follows, I first describe the methodology of the testing and then the results of my investigations. The chapter will conclude with an integral, empirically examined model of the semantic-syntactic characteristics of Romance collection nouns in present-day language.

4.1 Methodology

The empirical method chosen to examine my research questions and hypotheses consists of acceptability judgement tests. These have an important advantage over corpus analyses in that they can give insights on the use of low-frequency lexemes in constructions that are not in common use, and they have an advantage over the introspective approach in that they consider more than one (non-expert) speaker. The former aspect can be further exemplified by the following example: as described in chap. 3.1, a common test of the accessibility of an OMN’s internal plurality is the combination with distributive quantifiers like *various*. First of all, a possible OMN like Sp. *mobiliario* ‘furniture’ has a much lower usage frequency than its equivalent

plural object noun *muebles* ‘pieces of furniture’. The relative frequency in the web corpus *esTenTen11* of *mobiliario* in the singular is 7.02 per million, whereas *muebles* has a relative frequency of 13.36 per mil. Additionally, there are only 6 occurrences of *varios mobiliarios* (in the plural, there are no occurrences in the singular) ‘various furniture’ compared to 246 occurrences of *varios muebles* ‘various pieces of furniture’; both constructions have a relative frequency at or near zero (0.00 and 0.02 respectively). As a consequence, corpus analyses may support the findings of a particular experiment, but only acceptability judgement tests can detect the fine-grained variations between different collection nouns and diverse languages. In these tests, I consider in addition to French also Italian, Spanish and Portuguese to test the second semantic-syntactic hypothesis of cross-linguistic variation.

To be able to compare Romance CCNs and OMNs, a list of possible OMNs must first be created. This type of methodological direction, i.e. the creation of an inventory of collection nouns from OMNs to CCNs and not vice versa, has been adopted because OMNs seem to be less common (in type frequency) than CCNs – this makes it easier to create a manageable list of nouns. Kiss/Pelletier/Stadtfeld (2014) e.g. only identify a proportion of 0.58% of OMNs in their Bochum English Countability Lexicon (accessible at <http://www.count-and-mass.org/>), which represents 80 exemplars of this category (of 13,804 noun-meaning pairs annotated). Since there has been comparatively little research done on OMNs in Romance languages (cf. chap. 3.1), I started by collecting all OMNs that have been mentioned in the literature on English OMNs (cf. e.g. Markman 1985; Barner/Snedeker 2005). I then identified Romance equivalents and generally broadened the search to uncountable collection nouns found in dictionaries.³⁰ To determine the status of these potential OMNs in French, Italian, Spanish and Portuguese I analysed them on a lexicographic basis and via corpus queries, which suffice for this first exploratory analysis. The empirical basis for this preliminary study is represented by the *TenTen* corpus family, a collection of web corpora, accessible at www.sketchengine.eu. These corpora were created with the web-crawler *Spiderling* and contain mostly contributions in forums, blogs, comments on newspaper articles and so forth. Because of their huge amount of data (the Spanish corpus *esTenTen11* e.g. consists of 11 billion tokens), these corpora are ideal for carrying out quantitative analyses on everyday language. But they have even more advantages over smaller corpora like the *Frantext* corpus (~ 70 mil. tokens) or the *Corpus del Español* (CDE; ~ 2 billion tokens): the *TenTen* corpora, being web-based, have the advantage of reflecting the

³⁰ Many modern dictionaries like the *Diccionario general de la lengua española* have the possibility to search via their provided software for e.g. nouns which only occur in the singular and which contain Sp. *conjunto* ‘collection/group’ in their definition.

heterogeneity of colloquial as well as formal speech – *Frantext* for instance mainly features items from French literature and is thus less likely to reflect spontaneous colloquial speech. Additionally, there is a *TenTen* corpus for each of the Romance languages being studied here with comparable token numbers. In contrast, other corpora vary in token size, captured period of time, diatopic varieties or text genres. The French corpus *Frantext* for example is restricted to texts that were written in France, whereas the Spanish CDE captures a great variety of dialects. Furthermore, some corpora reflect a high degree of linguistic heterogeneity for a rather short period of time (e.g. CORPES XXI); other corpora however concentrate on a systematic reflection of the diachronic development of a language, resulting in a more homogeneous text base (e.g. *Frantext*). In general, the fact that the variety of corpora are created with different goals and are not designed to directly compare different Romance languages complicates a systematic comparative approach. This does not necessarily mean that the different *TenTen* corpora have the exact same conditions and token sizes (cf. Table 4.1), but they were at least created with the same tool. There is one possible alternative to the *TenTen* corpus family, namely the C-Oral-Rom corpora. These were intentionally created to compare phenomena of spoken language in French, Spanish, Portuguese and Italian (cf. Cresti/Monéglija 2005, 1). Unfortunately, the relatively low token frequency (cf. Table 4.1) of these corpora makes it impossible to find any (or enough) OMN occurrences – there are for instance no occurrences for Fr. *coutellerie* ‘cutlery’ and only 16 of Sp. *ropa* ‘clothing’. Table 4.1 summarises these considerations; the overview is not intended to be an exhaustive list, but should simply illustrate my explanations.³¹

Table 4.1: Comparison of token quantities and text types of different Romance corpora focusing on present-day language (token quantities are rounded).

	<i>TenTen</i> ³²	C-Oral-Rom	Diverse	
FR	<i>frTenTen12</i> 11 bil.	300,000	– <i>Frantext</i> : written language (lit.); FRA; 1960–2019 – CEFC: written/spoken language FRA/CH/BEL; 21st cent.	70 mil. ³³ 10 mil.

³¹ Annette Gerstenberg gives an excellent overview of the existing Romance corpora and text databases (<https://wikis.fu-berlin.de/display/tdr/Korpora+und+Textdatenbanken>, last accessed: March 25th, 2022).

³² After the analyses were finished in 2020, there was an update of the *TenTen* corpora, so that there is now additionally a *frTenTen17* (7 bil. token) and *esTenTen18* corpus (20 bil. token). These could not be taken into consideration afterwards.

³³ This token quantity refers to the chosen sub-corpus for present-day language use (since 1960). There is a combined token frequency of 291,058,214 (1120–2019) over the entire *Frantext* corpus.

Table 4.1 (continued)

	<i>TenTen</i>	C-Oral-Rom	Diverse	
IT	<i>itTenTen16</i> 6 bil.	300,000	– CORIS/CODIS: written/spoken language; 1980s/90s – CLIPS: written/spoken language; 1990s	130 mil. 3,8 mil
SP	<i>esTenTen11</i> 11 bil.	300,000	– CREA: written/spoken language; entire Spanish speaking world; 1975–2004 – CORPES XXI: written/spoken language; entire Spanish speaking world; 21st cent. – CDE (web): written language; entire Spanish speaking world; 2013–2014	154 mil. 300 mil. 2 bil.
PT	<i>ptTenTen11</i> 5 bil.	300,000	– CRPC: written/spoken language; various varieties – CDP: written/spoken language; various varieties; 2012–2019	311 mil. 1 bil.

To determine the status of a potential OMN in the different Romance languages, the two minimal definitional criteria of external unboundedness and internal plurality were applied (cf. chap. 3.1). To test the degree of unboundedness of a certain noun, I analysed via corpus queries in the respective *TenTen* corpus if the noun occurs mostly in the singular or the plural (or whether there are no tendencies) and whether it occurs together with the indefinite article or not. Both aspects are typically correlated with (un)countability in number marking languages (cf. Chierchia 2010, 109–111; Rothstein 2017, 75). Consequently, I searched each potential OMN in the inflectional singular and the plural form, as well as the construction <indef. art. + noun>³⁴ in the respective *TenTen* corpus. This latter proportion is calculated on the basis of all occurrences and not only the singular frequencies, because there are also cases of potential POMNs occurring together with an indefinite article – those are rare but they do exist. Those cases like in (36a) are mostly discursively marked, but unmarked examples like in (36b) can also be found. (36a) is marked because it is a contribution in a fantasy-medieval role play forum, so it can be assumed that the speaker is trying to imitate an archaic style and the example does not reflect actual language use. Example (36b) however does not represent a marked context, but

³⁴ The tagging of the different *TenTen* corpora only allows them to search for the tag *article* in general, and there is no possible way to automatically differentiate between indefinite and definite articles. Consequently, I searched manually for the construction: [word=*indef. art.*] [lemma=*noun*], e.g.: [word=*una*] [lemma=*ropa*].

simply reflects spoken language use. The example represents a contribution in a forum of the online role play game *World of Warcraft*, which is not necessarily characterised by medieval, but only fantasy-related features. In addition, the speaker in (36a) in his or her role is in contrast to example (36b), where the speaker addresses her statements directly to her fellow players (outside the role playing). This latter speaker furthermore presents herself as Camille, 17 years old, from Lyon – so possible language influence from Spanish, where the construction *una gente* ‘a person’ is common in Mexico (cf. DPD, s.v. *gente*; Real Academia Española/Asociación de Academias de la Lengua Española 2009, 803), is thus improbable.

- (36) a. Contre toute attente l’homme sortit de sa bourse *une pierreries*, attirant leur attention, dévoilant leur sourire à l’appétit de ce gain, leur jaune sourire beuh. . . (*frTenTen12*)
 ‘Contrary to all expectations, the man pulled out of his purse *a gemstone*, attracting their attention, revealing their smile of appetite to this gain, their yellow smile beuh. . .’
- b. C’est une bonne idée mais ce serait des groupes, des vrais groupes tout le temps ou pas? (par exemple quand on invite *un gens* dans le groupe avec les portraits des persos et tout) (*frTenTen12*)
 ‘That’s a good idea, but would it be groups, real groups all the time or not? (for example, when you invite *a person* into the group with profile pictures of the characters and everything)’

Table 4.2 below presents the results of this quantitative corpus analysis. The first two numbers represent the percentage of all occurrences in the singular and the plural, the last number represents the percentage of occurrences of the construction <indef. art. + noun>. Relative frequencies are not indicated in this context, to capture the proportions which are of greater relevance here. As two additional points of reference for the overall comparison, the results for English and German are represented too, since they served as a starting point for this initial exploratory analysis. I include in this table only those examples where there is a potential OMN in at least one of the Romance languages being studied here. The asterisk marks those cases where the potential OMN is polysemous between a collection noun reading and another, countable, interpretation. I will explain these in more detail in the following.

As can be seen in Table 4.2, it seems to be the case that there are more OMNs in German and English than in the Romance languages, where many of the equivalent concepts are expressed not by a heterogeneous mass noun but by a count

Table 4.2: List of potential Romance OMNs (SG/PL – percentage of occurrences with the indef. art.).

	French (frTen12)	Italian (itTen16)	Spanish (esTen11)	Portuguese (ptTen11)	English (enTen15)	German (deTen13)
ARTEFACTS						
<i>mobiliar</i>	94/6-7	<i>mobilia</i> 72/28-0	<i>mobiliario</i> 97/3-2	<i>mobiliário</i> 92/8-1	<i>furniture</i> 100/0-1	<i>Mobiliar</i> 100/0-0
		<i>arredamento</i> 83/17-4	<i>moblaje</i> 96/4-5	<i>mobilia</i> 80/20-2		
		<i>mueblaje</i> 95/5-0				
<i>bijouterie*</i>	89/11-15	<i>gioielleria*</i> 85/15-8	/	<i>bijuteria*</i> 82/18-0	<i>jewellery</i> 100/0-2	<i>Schmuck</i> 100/0-0
<i>pierreries</i>	2/98-0	<i>bigiotteria*</i> 98/2-0				
<i>courrier*</i>	85/15-13	<i>corriere*</i> 95/5-1	<i>correo*</i> 83/17-7	<i>correio*</i> 66/34-0	<i>mail*</i> 97/3-3	<i>Post</i> 100/0-0
<i>vaisselle</i>	97/3-1	<i>stoviglie</i> 1/99-0	<i>vajilla</i> 72/24-5	<i>louça*</i> 64/36-1	<i>kitchenware</i> 98/2-1	<i>Geschirr*</i> 96/4-2
		<i>vasellame</i> 96/4-0			<i>crockery</i> 100/0-1	
<i>habillement</i>	95/5-2	<i>abbigliamento</i> 99/1-4	<i>ropa</i> 91/9-0	<i>roupa*</i> 45/55-3	<i>clothing</i> 100/0-1	<i>Kleidung</i> 100/0-0
<i>vêtement*</i>	18/82-3				<i>clothes</i> 0/100-1	
<i>fringues</i>	0/100-0					
<i>linge</i>	94/6-9	<i>biancheria</i> 99/1-0	<i>lenceria*</i> 99/1-2	<i>lençaria*</i> 100/0-0	<i>underwear</i> 100/0-0	<i>Unterwäsche</i> 100/0-0
<i>verrerie*</i>	77/23-4	<i>vetrame</i> 46/54-0	<i>cristeria*</i> 80/20-4	/	<i>glassware</i> 99/1-0	<i>Glaswaren</i> 6/94-0
<i>mercerie*</i>	91/9-9	<i>merceria*</i> 70/30-9	<i>merceria*</i> 81/19-17	/	<i>haberdashery*</i> 96/4-9	<i>Kurzwaren</i> 2/98-0
<i>argent</i>	100/0-0	<i>denaro</i> 94/6-0	<i>dinero</i> 97/3-1	<i>dinheiro</i> 99/1-1	<i>money</i> 100/0-0	<i>Geld</i> 96/4-0
		<i>soldi</i> 4/96-0				
ANIMALS						
<i>feuillage</i>	72/28-4	<i>fogliame</i> 94/6-2	<i>folhaje</i> 91/9-2	<i>folhagem</i> 100/0-3	<i>foliage</i> 100/0-0	<i>Blattwerk</i> 99/1-0
<i>bétail</i>	98/2-1	<i>bestiame</i> 98/2-0	<i>ganado</i> 89/11-1	<i>gado</i> 98/2-1	<i>cattle</i> 100/0-2	<i>Vieh*</i> 100/0-1
<i>volaille</i>	56/44-3	<i>pollame</i> 98/2-0	<i>volateria</i> 91/9-0	/	<i>poultry</i> 100/0-2	<i>Geflügel</i> 100/0-0
<i>gibier</i>	88/12-5	<i>cacciagione</i> 98/2-0	<i>caza*</i> 96/4-1	<i>caça*</i> 79/21-1	<i>game*</i> 63/37-4	<i>Wild</i> 100/0-1
		<i>selvaggina</i> 100/0-0				
PEOPLE						
<i>gens</i>	0/100-0	<i>gente</i> 95/5-0	<i>gente</i> 98/2-0	<i>gente*</i> 98/2-0	<i>people</i> 98/2-0	<i>Leute</i> 0/100-0
<i>police</i>	95/5-2	<i>polizia</i> 99/1-0	<i>polícia*</i> 80/20-2	<i>polícia*</i> 95/5-1	<i>police</i> 100/0-4	<i>Polizei</i> 100/0-0

noun. Engl. *footwear* and the German equivalent *Schuhwerk* e.g. mostly have count noun equivalents in the Romance languages being studied (Fr. *chaussure*, Sp. *calzado*, Pt. *calçado* ‘shoe’). Only Italian has a potential OMN in this conceptual domain, viz. *calzatura*. It may of course also be the case that the basic direction of search for OMNs (German/English > Romance languages) has biased the range of potential nouns to be examined further.

The English and German potential OMNs show clear tendencies to be either used only in the singular (e.g. Engl. *crockery*) or in the plural (e.g. Germ. *Kurzwaren*). The general proportions of occurrences with the indefinite article are very low, and the existing ones mostly represent compounds with the particular noun as a modifier. For Engl. *furniture* there are e.g. 5,638 occurrences of *a furniture* (rel. frequency: 0.31), but these are mostly examples like *a furniture store*, *a furniture dealer*, *a furniture business* etc. The general proportion of the occurrences with the indefinite article in German is lower since compounds are graphically linked in this language, so it is consequently not possible to find occurrences like *ein Mobiliar Geschäft* ‘a furniture store’ but only *ein Mobiliargeschäft*. I consequently consider the data as relatively reliable – provided that we assume a smaller proportion of the indefinite article with OMNs in English. With regard to the Romance languages examined here the proportions are slightly less clear. Here there are many potential OMNs that are found relatively often in the plural and combined with the indefinite article. In many cases, however, these examples represent polysemous nouns that also have a count noun reading, like Fr. *bijouterie* ‘jewellery_{mass}/jeweller’s shop_{count}’. Therefore, the mere analysis of number frequencies and occurrences with the indefinite article does not only represent the OMN interpretation (‘jewellery’), but also the connected count noun interpretation (‘jeweller’s shop’) – these polysemous nouns are marked by an asterisk in Table 4.2.³⁵ For e.g. Sp. *cristalería* ‘glassware’ the quantitative analysis indicates an OMN-potential: the singular plural proportion is 80% to 20% and the lexeme occurs with the indefinite article in only 4% of occurrences. The *Diccionario general de la lengua española* (DGL, s.v. *cristalería*), however, lists both countable and uncountable uses and does not make any comments on the general countability of the noun in its use as a collection noun:

³⁵ For the lexicographic analysis I consulted the following dictionaries: The *Grand Robert de la langue française* (GR) for French, for Spanish the *Diccionario de uso del español de América y de España* (DGL), for Italian the *Zingarelli* and for Portuguese the *Novo Dicionário da Língua Portuguesa* (NDLP).

- (37) a. Establecimiento donde se fabrican o se venden objetos de cristal.
 ‘(Work)shop where glass objects are manufactured or sold.’
 b. Conjunto de objetos de cristal que forman parte de una vajilla.
 ‘Set of glass objects that are part of the tableware.’

Consequently, the actual status of Sp. *cristería* cannot be determined without a time-intensive qualitative analysis of the corpus data. The categorisation of Fr. *vaisselle* ‘crockery/dishes’ however is easier. Here the *Grand Robert* (GR, s.v. *vaisselle*) gives the following definitions:

- (38) a. Ensemble des récipients qui servent à manger, à présenter la nourriture.
 ‘All the containers that serve to eat, to present food.’
 b. Ensemble des plats, assiettes, ustensiles de table, etc., qui sont à laver.
 ‘All the dishes, plates, table utensils etc, that need to be cleaned.’
 c. Ensemble d’ustensiles, de récipients servant à un autre usage que la table.
 ‘All the utensils, containers that serve for other purposes than at the table.’

The *Grand Robert* also does not make any comments on the countability of the noun in question, but since *vaisselle* has no clear count noun senses as in e.g. ‘(work)shop’ and since the quantitative analysis confirms the status of a mass noun (almost all are singular occurrences only, and hardly any are used with the indefinite article), one can in fact determine a potential OMN status for this noun. Following on from these considerations, selection of potential OMNs in the Romance languages is still largely based on lexicographic information in combination with the corpus data summarised in Table 4.2.

Following the pre-selection of OMNs and against the background of the theoretical preliminaries described in part I of this present work, I have chosen the following objects of examination for the acceptability judgement study: according to the animacy hierarchy (cf. chap. 1.1) and its consequences on the linguistic behaviour of collections (cf. chap. 3.1), I determined three semantic domains of testing: inanimate, animate and human collections. Since the main goal of the study is a systematic comparison between different types of collection nouns and various Romance languages, I chose three OMNs that all have an equivalent CCN, as well as an equivalent object noun (ON) in all four Romance languages being studied. Depending on the test context, the latter noun type either represents a superordinate (like Fr. *animal* ‘animal’) or a subordinate (like Fr. *vache* ‘cow’) term and it is used either in the singular or the plural. The tested nouns are set out in the following Table 4.3:

Table 4.3: Romance OMNs and their quasi-synonyms in the acceptability judgement study.

Semantic domain	Collection noun sub-type	French	Italian	Spanish	Portuguese
CLOTHING	SOMN ‘clothing’	<i>vêtement</i>	<i>abbigliamento</i>	<i>ropa</i>	<i>roupa</i>
	POMN ‘clothes’	<i>fringues</i>	/	/	/
	ON ‘piece of clothing/ shirt’	<i>vêtement/ chemise</i>	<i>vestito/ camicia</i>	<i>prenda/ camisa</i>	<i>peça de roupa/camisa</i>
	CCN ‘outfit’	<i>tendue</i>	<i>outfit</i>	<i>atuendo</i>	<i>conjunto</i>
CATTLE	OMN ‘cattle’	<i>bétail</i>	<i>bestiame</i>	<i>ganado</i>	<i>gado</i>
	ON ‘animal/ cow’	<i>bête/ vache</i>	<i>bestia/ mucca</i>	<i>animal/ vaca</i>	<i>animal/ vaca</i>
	CCN ‘herd’	<i>troupeau</i>	<i>gregge</i>	<i>rebaño</i>	<i>rebanho</i>
PEOPLE	OMN ‘people’	<i>gens</i>	<i>gente</i>	<i>gente</i>	<i>gente</i>
	ON ‘person/ women’	<i>personne/ femme</i>	<i>persona/ donna</i>	<i>persona/ mujer</i>	<i>pessoa/ mulher</i>
	CCN ‘team’	<i>équipe</i>	<i>squadra</i>	<i>equipo</i>	<i>equipa/e</i>

It would be desirable to include a second set of artefact nouns to analyse possible differences with regard to the qualities of the constituting entities (e.g. heterogeneity, size, functionality). For the different reasons presented above, this would only be the case for the semantic domain of FURNITURE, but here many of the Romance equivalents seem to oscillate between an interpretation as an OMN and a CCN. The *Grand Robert* e.g. indicates for Fr. *meublier* ‘furniture’ a definition as a CCN “ensemble des meubles destinés à l’usage et à l’aménagement d’une habitation” (‘collection of pieces of furniture for the use and the fitting of a house’), since clearly there is the implication of spatio-temporal bounding. The *Grand Robert*, however, also gives the example *acheter du mobilier* ‘to buy furniture’, where the mass character predominates and a certain intensional stability is necessary to convey sense (cf. GR, s.v. *mobilier*). Depending on the context the semantic characteristics are thus more or less stable and the noun in question either manifests itself as a mass or as a count noun. In English, this opposition is perhaps best described by the distinction between *furniture* and *fittings*. A strict separation or selection of one or the other sense would be difficult to achieve, so I left this set out (but see the description of the follow-up study described in chaps. 4.2.1 and 4.2.2). A further comment is necessary with respect to the semantic domain of CLOTHING in French as well as in Spanish. In French, the prototypical OMN here would be *habillement* ‘clothing’, which shows clear preferences for the singular and which cannot be combined with the indefinite article (cf. Table 4.2). Its problematic property, however, is that this noun is generally not in

common usage. Its relative frequency in *frTenTen12* is only 3.02 in the singular in comparison to the 9.69 of *vêtement* ‘clothing/piece of clothing’, and it generally appears on websites like *wikisource.org*, *revue.org* or *forumactif.com* where language use from earlier phases or its imitation has to be assumed. In particular, the domain of *forumactif.com* hosts a lot of role play forums where the users imitate the style of the Middle Ages in collaboratively telling stories of witches, dragons and castles (e.g. *The Knights of Emerald*). Consequently, the results of an acceptability judgement study including *habillement* would be distorted, not necessarily because of its unacceptability in a certain construction, but simply because of its unusualness. For these reasons, I chose not the SOMN *habillement*, but the POMN *fringues* ‘clothes’. This too is not the unmarked option – this would always be the flexible *vêtement* – because it is generally marked as popular or colloquial (cf. GR, s.v. *fringues*), but it is the best compromise in this case. I also tested the common variant *vêtement* in possible mass syntax contexts. Any deviations or conspicuous features with respect to this French variant will be addressed in the respective analyses. In Spanish, the default equivalent for ‘clothing’ is *ropa*. As shown in Table 4.2, this is predominantly used in the singular and is not in practice combined with the indefinite article (the exact proportion is 0.23%). These corpus data therefore lead to the noun’s classification as a SOMN. There is, however, no uniform lexicographic definition of this noun confirming the assumed status and hardly any dictionary comments on the countability status of *ropa*. The *Diccionario general de la lengua española* defines it as “nombre genérico de cualquier pieza de tela confeccionada que viste a una persona, un objeto o un lugar” (‘generic name for any piece of made-up fabric that dresses a person, an object or a place’) (DGL, s.v. *ropa*), thus focusing on its status as a superordinate without specifying whether *ropa* denotes individual entities or a collection. In contrast, the *Diccionario de la lengua española* simply defines it as “prenda de vestir” ‘piece of clothing’ without referring to the collection noun uses of *ropa* (DRAE, s.v. *ropa*). One of the only dictionaries that acknowledges the ambiguous status of *ropa* is the *Diccionario de uso del español*: “Referiéndose a los vestidos, puede usarse en singular o en plural” (‘when referring to clothes, it may be used in the singular or the plural’) (MM, s.v. *ropa*). Sp. *ropa* thus seems to be ambiguous: although it syntactically behaves like a SOMN, it may also be used as a countable object noun. A similar problematic issue relates to the Spanish, Portuguese and Italian SOMN *gente* ‘people’ and the French POMN *gens* ‘people’. It is simply not possible to undertake a systematic comparison of singular and plural OMNs comprising the same kind of referents in the same Romance languages. While we can say that It., Pt. and Sp. *gente* and Fr. *gens* are good candidates for this comparison – all four OMNs designate an indefinite mass of people – other language-specific aspects may come into play here. For example, the different

degrees of marking the mass-count distinction as summarised in Table 1.2 or the morphological singular or plural (cf. Figure 3.1) may influence accessibility of the internal plurality. For these reasons, the four variants will be analysed together, but their results will be interpreted separately, at least to a certain extent. In this respect, a special note on Pt. *gente* is necessary: In the Brazilian variety *a gente* '(lit.) the people' is grammaticalised as a personal pronoun of the first person plural, increasingly replacing the older pronoun *nós* 'we' (cf. Lopes 2002). For *gente* not being exclusively tested with the definite article *a*, but also with distributive quantifiers etc, I nevertheless include it in this study. It should be born in mind that possibly negative evaluations of this lexeme might not necessarily be due to its status as a SOMN, but because it is no longer considered a fully fledged noun. In the European variety, this pronominalisation process is not apparent (cf. Vianna 2011). Finally, I chose the respective Romance equivalents of 'women' and 'person' as object noun control variants and not of 'man' to avoid the inherent ambiguity of Fr. *homme*, It. *uomo*, Sp. *hombre* and Pt. *homem* between 'man' and 'human'. The latter sense may blur the comparability between this count noun and the respective collection nouns, since it is a typical count superordinate – in contrast to the tested OMNs – that tends to be used in the plural (cf. Wisniewski/Murphy 1989; Mihatsch 2006, 144).

The acceptability judgement test examined the delimiting semantic-syntactic domains of collection nouns of external boundedness and internal plurality. Following on from this, a total of five different tests with respect to these domains, whose exact structure will be explained in the respective sub-chapters, was conducted:³⁶

1. External boundedness
 - a. Compatibility with the indefinite article
 - b. Compatibility with distributive quantifiers
2. Internal plurality
 - a. Compatibility with stubbornly distributive predicates
 - b. Compatibility with highly distributive predicates like *one after the other*
 - c. Hierarchical relations structuring the collection

³⁶ I am indebted to all the native speakers who supported me in discussing the test sentences of the experiments in the different Romance languages. In particular, I thank Benjamin Massot and Nicolas Heslaut for French, Sarah Dessi Schmid, Valentina Vincis and Vittorio Prada for Italian, María José Gallucci, Ana Vazeilles, Carla Miotto and María Xesús Bello Rivas for Spanish as well as Ina Biscaia Berner, Priscilla Nogueira and Suzana Vasconcelos de Melo for Portuguese. Without their support and endless patience, this study could not have been carried out.

Tests 1a and 1b address the second semantic-syntactic research question (RQ_{ss}2) and the assumption that there is a Romance continuum of syntactic flexibility of OMNs, i.e. their combinability with the indefinite article as well as their combinability with the inflectional plural and more or less distributive quantifiers. Tests 2a and 2b address the first and third semantic-syntactic research questions (RQ_{ss}1/3) in examining accessibility of the internal plurality of a collection noun depending on its external boundedness and the ontological type of referents. Test 2c addresses the hierarchical relations between constituting entities of a collection and the collection itself.

Before coming to the actual description of the test design, participants and results, an important note on the overall methodology is necessary. As represented by the rather high number of different tests examining the various types of collection noun and the quasi-synonyms shown in Table 4.3, the present study aims to give a full picture of many of the semantic-syntactic characteristics elaborated in chap. 3.1. This procedure has a number of advantages and disadvantages. The main advantage lies in the fact that the study does not only examine one single type of collection noun, nor does it examine only one or merely a few semantic or syntactic characteristics, it thus allows portrayal of a broader picture of various lexemes and their multiple properties. Many studies conducted until now have, for example, mainly focused on the Engl. SOMN *furniture* with the consequence that they are even called *furniture-nouns* (cf. e.g. Grimm 2012). The main disadvantage of this procedure is, however, that a straightforward Latin-square design, which tests a few conditions with various linguistic representations, is not possible. This is because I assume significant differences between various types of the same kind of collection noun in accordance with the animacy hierarchy. I therefore cannot automatically view them as interchangeable representations of the same type. A subsuming operation of various CCNs, for instance – at least at this point of the analysis – would be inappropriate and I would still need to regard them as exemplars for a hypothetical category. Importantly, for the statistical analyses that I will conduct on the test results, this entails that I cannot use e.g. an analysis of variance with repeated measurements, since every participant in fact only rated one instance for each condition (e.g. the compatibility of the Fr. inanimate POMN with a stubbornly distributive predicate, i.e. *fringues longues* ‘long clothes’, but not the animate equivalent *bétail maigre* ‘lean cattle’). This reduces the reliability of the data, but also opens up the possibility of detecting idiosyncrasies and noun-specific characteristics. This is a crucial point because, as said before, research on OMNs has mainly been restricted to *furniture*.

The acceptability judgement study was conducted online (realised with OnExp: <https://onexp.textstrukturen.uni-goettingen.de/>) and to a certain extent also on paper. The latter version was necessary since French and Portuguese speakers' response to the online participant acquisition was not sufficient for our needs so that the data collection had to be made additionally on the spot, namely in Paris and Lisbon.³⁷ In both versions of the test, the participants rated a total of 54 pseudo-randomised sentences on a Likert scale of 1 'not acceptable' to 7 'perfectly acceptable'. Of these 54 sentences, about 17 or 18 were actually testing the properties of Romance collection nouns, and the remaining 36 to 37 sentences were fillers consisting of test sentences relating to a parallel research project on verbal aspectuality (24 sentences)³⁸ and 'real' fillers not related to the objects of analysis (11 to 12 sentences).³⁹ To match the test sentences, the latter were deliberately modified in such a way that they represented very good sentences, rather weird ones and essentially unacceptable ones at the levels both of syntax and of semantics.⁴⁰ In the online version of the study, the participants saw the sentences one at a time, but in the paper version this was not possible, so in this case sentences were presented beneath each other. There are no significant differences between the two versions of the French and the Portuguese study. The French sentences, taken together all test and control sentences of the nominal domain, were rated online with an overall average of 4.88 ($n_{\text{sentence}} = 1,143$) and on paper with an average of 4.86 ($n_{\text{sentence}} = 692$). The difference of 0.02 is statistically not significant ($t(1,833) = .28, p = .783$). The Portuguese sentences of the nominal domain were rated online with an average of 4.71 ($n_{\text{sentence}} = 491$) and on paper with 4.46 ($n_{\text{sentence}} = 1,315$). The difference of 0.25 between these two mean ratings is statistically significant ($t(834.32) = 2.29, p = .023, d = .01$), but the effect size is negligible.⁴¹ The two versions of French and Portuguese are thus treated as one.

37 I would like to thank Charlotte Coy and Ina Biscaia Berner who conducted the study on-site. I also heartily want to thank Sarah Dessi Schmid, Wiltrud Mihatsch, Ana Vazeilles, Maria José Gallucci, Vassil Mostrov, Mathilde Huguin, Roberta Pires de Oliveira and many others who helped spread the study online.

38 See Dessi Schmid 2021 for the results on Romance progressive verbal periphrases in combination with states.

39 Since the overall study was a collaborative project within the framework of the CRC833 project on verbal and nominal aspectuality between lexicon and grammar, I will in what follows use the personal pronoun *we* when referring to the study in total and *I* when describing my own specific analyses and calculations.

40 See <https://hdl.handle.net/11022/0000-0007-EB93-A> for a list of sentences used.

41 The acceptability ratings are considered not as necessarily ordinal, but as nominal scales. This methodological decision was made since I am mainly interested in the differences between the control condition and the various test conditions (e.g. CCNs vs. OMNs) and not the absolute

The Spanish, the Italian and the Brazilian versions were exclusively conducted online.

At the beginning of the test, participants saw the example rating displayed in Figure 4.1; there were no practice items, and the examples were assumed to suffice as anchor items.

Aujourd'hui, j'ai téléphoné à ma mère pendant deux heures.

C'est une phrase normale et quotidienne, vous pouvez l'évaluer avec 6 ou 7.

Aujourd'hui, j'ai écrit des lettres à ma mère en deux heures.

Cette phrase est un peu bizarre, vous pouvez l'évaluer avec 4.

J'ai mangé en deux heures ma mère aujourd'hui.

Cette phrase ne représente pas l'usage normal, vous pouvez l'évaluer avec 1 ou 2.

Figure 4.1: Example rating in the acceptability judgement study.

Table 4.4 shows an overview of the participants for each language. The displayed quantities represent the already corrected numbers, since some participants had to be excluded from the analysis, either because they did not finish the study or because they rated two or more control sentences lower than the double standard deviation. In this case, I must assume that they did not have the same interpretation of the rating scale or level of concentration as the other participants. Thirdly, due to technical reasons, the demographic data for some participants was not transmitted so that I had to exclude these as well. After this sorting of the participants, the following subjects were considered for the acceptability judgement study's evaluation:

Table 4.4: Overview of the participants of the acceptability judgement study.

	French		Italian		Spanish		European Portuguese		Brazilian Portuguese	
N	107		147		157		110		90	
Sex	♂	35	♂	54	♂	50	♂	36	♂	40
	♀	72	♀	90	♀	107	♀	76	♀	49
	♀	0	♀	3	♀	0	♀	0	♀	1

rating on the scale. I am aware of the fact that some informants generally tend to rate sentences lower and others higher, but by always sticking to relation analysis, these effects are weakened. It also enables me to perform these standard parametric tests like t-tests and ANOVAs (cf. Schütze/Sprouse 2013, 33).

Table 4.4 (continued)

	French		Italian		Spanish		European Portuguese		Brazilian Portuguese	
Age	18–29	89	18–29	78	18–29	61	18–29	68	18–29	46
	30–49	13	30–49	53	30–49	79	30–49	26	30–49	37
	50+	5	50+	16	50+	17	50+	16	50+	7
Origin	FR	7	NoIT	63	CA	62	M/SPO	85	SBR	64
	FrEu	12	C/SIT	63	SA	38	NoPO	8	NoBR	7
	FrAfr	1			EU	43				
	other	10	other	20	other	4	other	14	other	17
	NA	5	NA	1	NA	10	NA	3	NA	2
Lev. Ed.	no	1	no	1	no	0	no	1	no	0
	VcEd./	37	VcEd.	37	VcEd./	10	VcEd./	66	VcEd./	5
	Grad.		Grad.		Grad.		Grad.	41	Grad.	
	Univ.	64	Univ.	107	Univ.	147	Univ.		Univ.	85
	NA	5	NA	2	NA	0	NA	2	NA	0

We consider the European and Brazilian varieties of Portuguese as distinct categories in the study for two reasons. On the one hand, the lexical and grammatical differences between the two diatopic varieties made it necessary to create two versions of the experiment in the first place. On the other hand, as described in chap. 1.2.2, I assume important differences between European and Brazilian Portuguese with regard to their coding of the mass-count distinction and consequently the significant differences between them in terms of the acceptability of OMNs in certain contexts. As can be seen in Table 4.4, irrespective of the language tested, most participants were women of between 18 and 49 years of age. This age span might be due to participant acquisition mainly having been conducted by means of university mailing lists and similar channels of communication, as most of the participants are presumably students or working at a university. This is reflected by the level of education indicated, which for the most part indicates an already achieved academic degree (Univ.), a vocational education or being at the undergraduate level (VcEd./Grad.). The French participants come mostly from France (FR), but also from other French speaking European countries (FrEu), French-speaking Africa is only marginally represented (FrAfr). Italian and Spanish participants are equally divided into the main dialectal regions: Southern and Central Italian (C/SIT) or Northern Italian (NoIT), as well as Central America (CA; mostly Venezuela), Southern America (SA; mostly Argentina) or Europe (EU). Due to the necessary on-site testing in Portugal, most participants come from Lisbon and surrounding areas. The Brazilian participants primarily come from the southern coast

regions (mainly São Paulo, Rio de Janeiro and Florianópolis). Some participants did not indicate their origin (NA) and some also come from Germany or other non-Romance countries (other). Considering all ratings regarding the testing of nominal aspectuality (test and control sentences), there are basically no statistically significant differences between the different demographic groups. With regard to sex, the three Italian participants who indicated *other* in this category could not be considered because of the very small quantity in relation to the other two categories. There are no statistically significant differences between the mean ratings of the single sentences by the male and the female participants in any of the four languages.⁴² There is a tendency for older people to rate the sentences lower than the age groups of younger and middle-aged participants. This is, however, a general tendency and not dependent on a certain test condition, i.e. even if there is a significant difference between the three age groups, there is never a significant interaction with the test condition.⁴³ With regard to participants' different regions, I did not calculate any comparisons for French and Portuguese (European and Brazilian), since most participants come from a single (dialectal) region (cf. Table 4.4). For Italian, I only considered the two main regions of Northern Italy vs. Central and Southern Italy. This distinction is based on the general assumption that the mass-count distinction is marked to a lesser degree in Northern Italian dialects than in the Central and Southern varieties: as described in chap. 1.2.2, the latter varieties have an obligatory partitive article as well as the *neutro de materia*, and they thus mark the mass-count distinction to a greater degree, i.e. more overtly and with a higher degree of obligatoriness (cf. Carlier/Lamiroy 2014, 506–514; Loporcaro 2016, 933–934; Pomino 2017, 695–697). These dialect-specific traits may influence the test results in displaying the same tendencies as between e.g. European and Brazilian Portuguese: a lesser degree of grammaticalisation of the mass-count distinction may correlate with a higher saliency of the constituting entities. To test whether there are significant differences between the two dialect groups in Italian, a two-way ANOVA with the factors *dialect group*

42 $t_{FR}(1,833) = 1.19, p = .236$ ($M_{\text{♂}}: 4.80, M_{\text{♀}}: 4.91$); $t_{IT}(2,425) = 1.44, p = .150$ ($M_{\text{♂}}: 4.63, M_{\text{♀}}: 4.76$); $t_{SP}(2,601) = .69, p = .489$ ($M_{\text{♂}}: 4.93, M_{\text{♀}}: 4.99$); $t_{EUP}(1,131.48) = 2.12, p = .034, d = .106$ ($M_{\text{♂}}: 4.68, M_{\text{♀}}: 4.46$); $t_{BRPT}(1,445) = 1.23, p = .219$ ($M_{\text{♂}}: 5.12, M_{\text{♀}}: 5.25$).

43 Two-way ANOVA *rating by age group*condition* (the factor *condition* will be treated in detail in the following chapters):

Fr.: $F_{\text{age}}(2, 1,823) = 5.08, p = .006, \eta_p^2 = .006$; $F_{\text{age*cond.}}(6, 1,823) = 1.46, p = .189$.

It.: $F_{\text{age}}(2, 2,460) = .03, p = .968$; $F_{\text{age*cond.}}(6, 2,460) = 1.84, p = .088$.

Sp.: $F_{\text{age}}(2, 2,594) = 3.12, p = .045, \eta_p^2 = .002$; $F_{\text{age*cond.}}(4, 2,594) = 1.03, p = .390$.

EuPt.: $F_{\text{age}}(2, 1,797) = 1.00, p = .368$; $F_{\text{age*cond.}}(4, 1,797) = .59, p = .671$.

BrPt.: $F_{\text{age}}(2, 1,454) = 4.49, p = .011, \eta_p^2 = .006$; $F_{\text{age*cond.}}(4, 1,454) = 1.55, p = .185$.

and *condition* was conducted, where the dependent variable was the rating (independent of the single tests). The ANOVA revealed no main effect on the dialect group ($F_{\text{dialect}}(1, 2,107) = .11, p = .735$) and a significant, but extremely weak interaction between the two main factors ($F_{\text{dialect} \times \text{cond}}(3, 2,107) = 4.57, p = .003, \eta_p^2 = .006$). With regard to Spanish, I did not compare the three main dialectal groups of Central and Southern America as well as Europe. There is the possibility that the weakening of the noun final /-s/, which marks the plural, in various regions of the Spanish-speaking world may have an influence on the marking of the mass-count distinction and thus the test results. This phonetic phenomenon is, however, highly heterogeneous and depends not only on geographical but also on social distribution (cf. Real Academia Española/Asociación de Academias de la Lengua Española 2011, 198–203). I will consequently not consider this further. I will therefore treat the results regarding possible demographic differences as homogeneous, irrespective of slight tendencies reported.

For different reasons, to be explained in the respective sections, a follow-up study was conducted which analysed various details of the main study on a slightly smaller scale. This follow-up study was exclusively conducted online via OnExp. The follow-up study consisted of 25–28 pseudo-randomised sentences which were presented to each participant one by one. In this test sequence, the participants saw 8–9 sentences testing nominal aspect and 17–19 filler sentences (10–12 again testing verbal aspectuality and 6–7 real filler sentences,⁴⁴ this time, all having a good acceptability). The participants were now given a single example of a good sentence (cf. (39)) to anchor the test items on the given scale (again from 1 ‘not acceptable’ to 7 ‘perfectly acceptable’ as in the main study), because – as will be shown by the main study – the reasons for lower evaluations were very heterogeneous and an exact anchoring was therefore difficult to obtain.

(39) Marie s’imagine: “Pour ma lune de miel, j’adorerais faire un voyage en Thaïlande.”

‘Marie imagines: “For my honeymoon, I would like to travel to Thailand.”’

The basic pieces of information relating to the participants of this follow-up study are summarised in Table 4.5.

⁴⁴ See <https://hdl.handle.net/11022/0000-0007-EB9F-E> for the sentences used.

⁴⁵ In the French and Brazilian Portuguese study, two participants each did not indicate their age.

Table 4.5: Overview of the participants of the follow-up acceptability judgement study.

	French		Italian		Spanish		European Portuguese		Brazilian Portuguese	
N	276		183		164		51		176	
Sex	♂	54	♂	47	♂	52	♂	8	♂	42
	♀	219	♀	135	♀	112	♀	41	♀	134
	♀	3	♀	1	♀	0	♀	2	♀	0
Age ⁴⁵	18–29	123	18–29	102	18–29	78	18–29	29	18–29	66
	30–49	121	30–49	59	30–49	66	30–49	19	30–49	85
	50+	33	50+	22	50+	20	50+	3	50+	23
Origin	FR	242	NoIT	64	CA	43	M/SPO	27	SBR	152
	FrEu	10	C/SIT	105	SA	48	NoPO	11	NoBR	10
	FrAfr	5			EU	55				
	other	18	other	10	other	13	other	12	other	11
	NA	1	NA	4	NA	5	NA	1	NA	3
Lev. Ed.	no	0	no	2	no	0	no	0	no	2
	VcEd./	35	VcEd./	70	VcEd./	37	VcEd./	22	VcEd./	4
	Grad.		Grad.		Grad.		Grad.	29	Grad.	
	Univ.	240	Univ.	111	Univ.	126	Univ.		Univ.	170
	NA	1	NA	0	NA	1	NA	0	NA	0

The demographic distribution is very similar to the main study: the participants were mainly women between the ages of 18 to 49 with a high level of education. Because the same channels of communication were used, it is at least possible that some participants also participated in the main study, but since there was a pause of about a year between the two test series, we do not consider this as potentially problematic. Just as in the main study, there are no significant differences between the evaluations of the male and female participants taking into consideration all sentences testing nominal aspectuality,⁴⁶ and are again there no significant differences between the three age groups in relation to the test conditions.⁴⁷ I also compared the evaluations of the two main dialectal groups

46 $t_{FR}(2,307) = 1.38, p = .167$ ($M_{\text{♂}}: 4.84, M_{\text{♀}}: 4.98$); $t_{IT}(1,636) = .19, p = .846$ ($M_{\text{♂}}: 4.72, M_{\text{♀}}: 4.69$); $t_{SP}(984.62) = 1.49, p = .137$ ($M_{\text{♂}}: 4.95, M_{\text{♀}}: 4.78$); $t_{EuPT}(396) = 1.89, p = .059$ ($M_{\text{♂}}: 5.02, M_{\text{♀}}: 5.46$); $t_{BrPT}(588.22) = 1.00, p = .318$ ($M_{\text{♂}}: 5.80, M_{\text{♀}}: 5.90$).

47 Two-way ANOVA rating by age group*condition (the factor condition will be treated in detail in the following chapters):

Fr.: $F_{\text{age}}(2, 2,304) = 4.22, p = .015, \eta_p^2 = .004$; $F_{\text{age*cond.}}(6, 2,304) = 1.21, p = .297$.

It.: $F_{\text{age}}(2, 1,635) = 1.30, p = .273$; $F_{\text{age*cond.}}(6, 1,635) = 1.84, p = .089$.

Sp.: $F_{\text{age}}(2, 1,454) = .42, p = .658$; $F_{\text{age*cond.}}(6, 1,454) = .68, p = .666$.

EuPt.: $F_{\text{age}}(2, 401) = .43, p = .651$; $F_{\text{age*cond.}}(6, 401) = 1.33, p = .243$.

BrPt.: $F_{\text{age}}(2, 1,522) = 4.17, p = .016, \eta_p^2 = .005$; $F_{\text{age*cond.}}(6, 1,522) = 1.52, p = .1167$.

of Italian. A two-way ANOVA with the two main factors *dialect* and *condition* revealed a significant, but very small main effect on dialect ($F(1, 1,513) = 20.18$, $p = .000$, $\eta_p^2 = .013$), as well as a significant, albeit small, interaction between the two main factors ($F(3, 1,513) = 11.63$, $p = .000$, $\eta_p^2 = .023$). Northern Italians tended to rate the test conditions lower (mean: 3.58) and the control condition higher (mean: 5.70) in comparison to the Southern Italians (mean: 4.54/5.30). Since the effect sizes are only small and there was no significant interaction between the condition and the diatopic varieties in the main study, I will not consider dialectal variation further.

In the following, I will present the various test domains in repeating firstly the central aspects of the state of the art summarised in chap. 3.1 as well as the hypotheses based on these aspects and, secondly, in describing the test results⁴⁸ and implications for and refinements of the integral semantic-syntactic model of collections as depicted in Figure 3.1. For each of the test sentences, I take the OMNs of Table 4.3 as a reference point. This is because research on CCNs has already determined in great detail many of the aspects addressed in this analysis. This type of collection noun is thus taken mostly as a point of comparison, but not specifically as a main object of examination. For each test, the exact selection of noun types will be described and explained in detail.

4.2 Results

4.2.1 Testing the external boundedness of count collective nouns and object mass nouns

As has been described in chap. 1.2, the aspectual property of boundedness corresponds to the possibility of linguistic operations such as counting. Only if an entity is construed as being externally bounded may its linguistic reflection be compatible with different syntactic means of expression of this boundedness. This includes compatibility with the indefinite article and the inflectional plural on the one hand as well as with numerals and other distributive quantifiers like *various* and the inflectional plural on the other. These features are tested by the first part of the acceptability judgement study, set out in the following. The hypothesis with regard to these tests, and based on the state of the art outlined

⁴⁸ The overall results as well as a list of all test sentences of the main acceptability judgement study are to be found in <https://hdl.handle.net/11022/0000-0007-EB96-7>. The results of the follow-up acceptability judgement study, as well as the test sentences used here are to be found in <https://hdl.handle.net/11022/0000-0007-EBA1-A>.

in chap. 3.1, is that CCNs are construed as bounded entities and that they should thus be easily compatible with the indefinite article. In contrast, OMNs, referring to unbounded entities, should not be compatible with this determiner. As we saw in chap. 3.1.1, this incompatibility is related to the fact that OMNs do not reflect a clearly determined number of entities – we simply do not know what would count as one entity. I expect SOMNs and POMNs to behave the same since they are both transnumeral, neither singular nor plural. In addition to this test, which is applicable to both morphologically singular and plural OMNs in all languages studied here, I present a test that examines the compatibility of this aspectual type with the inflectional plural and distributive quantifier *various* as well as small round numbers taking the two semantic domains of CLOTHING and PEOPLE as examples. The two different types of quantifier should show a varying degree of acceptability with the tested collection nouns: following Allan (1980, 548–549), *various* is a *fuzzy denominator* that does not specify the exact quantity of a plurality, but only estimates it. In contrast, round numbers specify the respective quantity to a higher degree. Following from this, *gens* and *fringues* should be more acceptable with the fuzzy denominator because it is more compatible with the indefiniteness of the collection. In opposition to this, POMNs should not be acceptable in combination with a round number since they semantically need to access an exact number of referents not available in this case – the exact denotation of *gens* and *fringues* is left unspecified. In addition, the effect may be increased by the use of a small number. Quantities of about two or three are perceptually grasped without counting – they are thus perceived as a holistic collection (cf. Hurford 1987, 93–95). This should not be possible with *gens* and *fringues* since their unboundedness hinders this effect of *subitising*, thus the perceiving a small number of referents holistically as a whole without the need to count them. The test sentences can hardly examine numerals this small and still be equivalent to the *various*-context, but we test the numeral *five* to at least come close to this.

As has been described above, tests 1a and 2b both will test the assumption of cross-linguistic variation of OMNs with respect to their syntactic mass noun properties. Specifically, I assume French OMNs not to be compatible with the indefinite article, but Brazilian Portuguese OMNs to be felicitous in this kind of construction, Italian, Spanish and European Portuguese OMNs may lie somewhere in between these two poles. In addition, I equally assume Brazilian Portuguese OMNs to be felicitous with the inflectional plural as well as various kinds of distributive quantifier, while Italian, Spanish and European Portuguese OMNs should not be felicitous in these kinds of constructions. This assumption is based on the degree of grammaticalisation of the mass-count distinction in these languages: whereas Spanish, Italian and European Portuguese still overtly mark mass syntax with bare singulars only in argument position, Brazilian Portuguese does not overtly mark mass syntax

(cf. Table 1.2). Assuming this, the former languages should be less flexible with the insertion of a syntactically mass noun in count syntax. Since the two tested OMNs in French are morphologically plural, I assume them to be as acceptable as the control condition in the case of a non-distributive plural, but to show a lesser degree of acceptability with the two kinds of distributive quantifiers, in accordance with the non-specificity of an entity that can be counted.

Test 1a: To test compatibility with the indefinite article, I tested the different OMNs featured in Table 4.3 and compared their acceptability in this construction with that of bounded entities like CCNs and singular object nouns (SONs). To create the test sentences I chose a syntactic context that does not allow for a sortal reading, since this is also possible for OMNs: *?J'ai acheté un mobilier* 'I bought a furniture' vs. *J'ai acheté un mobilier tout neuf* 'I bought a completely new furniture' (Flaux 1999, 484). Given the possibility that one of the OMNs may perhaps show a certain degree of acceptability with the indefinite article (cf. Table 4.2), it is additionally interesting to examine whether it is interpreted in these cases as either a bounded individual entity or an equally bounded collection. For this reason, I chose different verbs which either select the individual interpretation (e.g. *to vaccinate*) or the collective interpretation (e.g. *to drive together*). The action of vaccinating cattle focuses on the individual animals, whereas driving together cattle needs more than one referent to be acceptable. In accordance with these preliminaries, I tested the different OMNs of Table 4.3 in both distributive and collective contexts to examine their acceptability with the indefinite article. Because it is already clear that CCNs are perfectly fine in combination with the indefinite article, they are used here as control variants for the collective contexts, while the individual contexts have a SON as control variant. The semantic domain of PEOPLE is not tested in this analysis. As shown by Table 4.2, the various equivalents of 'people' simply do not appear in combination with the indefinite article in the corpora, whereas the other OMNs do show at least the possibility of such a construction. A further analysis of this combination would therefore not be very fruitful and is left out for this test. An example of the exact test sentences is given in (40) for French.

- (40) a. J'ai passé toute ma journée à faire le ménage, du coup il ne me reste qu'à <repasser_{DistrV} *une* fringue_{OMN}/*une* chemise_{SON}>.
'I spent the whole day doing housework, but now I just have to <iron a clothing/a shirt>.'
- b. Julie est très à la mode, elle sait que le plus petit accessoire peut <perfectionner_{CollV} *une* fringue_{OMN}/*un* ensemble_{CCN}>.
'Julie is very fashion-conscious, she knows that the smallest accessory can <perfect a clothing/an outfit>.'

- c. Marie fait un stage chez le vétérinaire, aujourd’hui elle apprend comment on <vacciner_{DistrV} un bétail_{OMN}/une vache_{SON}>.
 ‘Marie undertakes a work placement at a vet, today she learns how to <vaccinate a cattle/a cow>.’
- d. Marie fait un stage à la ferme, aujourd’hui sa tâche est de <rassembler_{CollV} un bétail_{OMN}/un troupeau_{CCN}>.
 ‘Marie undertakes a work placement at a farm, her task today is to <drive together a cattle/a herd>.’

It was not possible to test more than one verb of the two types since, as was mentioned above, I wanted rather to test more contexts and not just a few contexts with more variants. The test results are thus strictly speaking only exemplary. The results of this test 1a are displayed in Figure 4.2.

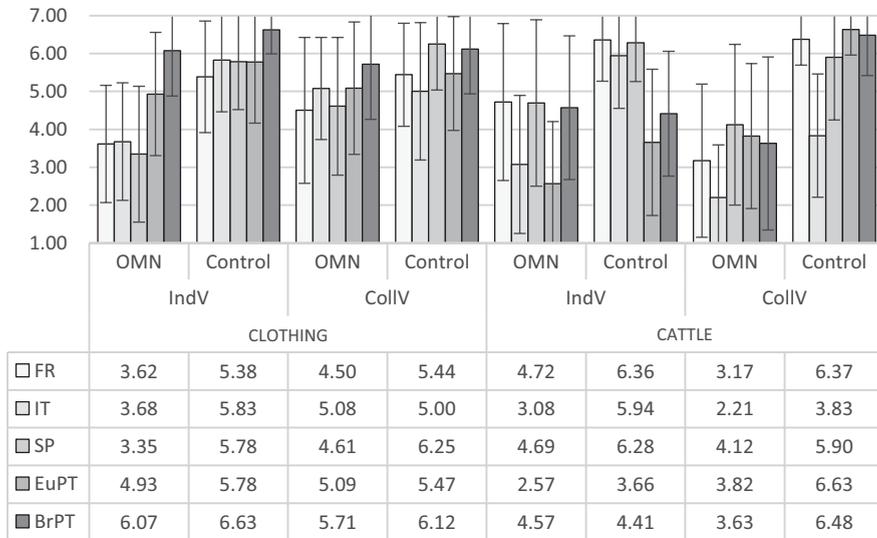


Figure 4.2: Mean acceptability ratings of Romance OMNs in combination with the indefinite article.

Figure 4.2 shows the mean acceptability rating of the different test and control sentences for each of the tested languages taken individually. Two observations are of importance here: on the one hand, taking only the mean ratings, all the OMNs tend to be generally rated lower than the mostly good control sentences. One can additionally observe the tendency for BrPt. *roupa* ‘clothing’ to be perfectly compatible with the indefinite article, just like the control variant. The OMNs in

the domain of CATTLE seem to function less well in this kind of context than the OMNs in the domain of CLOTHING, whose mean ratings are generally better. There seem to be no clear differences between the two verb variants tested. One may thus conclude that OMNs in the Romance languages are less compatible with the indefinite article than the respective control variant, and that the mean ratings of the animate domain are a little lower than of the inanimate domain. One may additionally summarise that the mean ratings are, however, not at the impossible end of the scale: a construction like *une fringue* is weird but not ungrammatical. On the other hand however, by taking into consideration the standard deviations for each sentence, there is the important observation that those are very high – especially in the case of the test sentences but also, to a slightly lesser degree, for the control sentences. This high degree of variation might be explained by the fact that neither the distributive nor the collective verb constructions were interpretable for the participants so that maybe some just rated the sentence low and others tried to coerce the meaning of it to give it some sense. There is, however, no possible point of comparison, i.e. a ‘neutral verb’, to examine this hypothesis any further. For this reason, the already mentioned follow-up study was conducted to repeat the testing under more controlled conditions.

The focus of this follow-up study lay mainly in levelling out the influence of specific verbs and specific contexts. In the main study, I was only able to test the compatibility with OMNs in a very limited number of sentences which did not allow me to generalise from them. I consequently repeated the testing of this construction, but increased the number of contexts tested. In particular, I now concentrated mainly on testing OMNs in the domain of CLOTHING. A frequency analysis of the *TenTen* corpora revealed that the different equivalents of ‘cattle’ simply do not seem to be very present in the daily life of speakers. They appear most often on government websites, on sites like *wikisource.org*, which imply historical texts, website of farmers’ associations and so on. I concluded from this analysis that test and control sentences of the various expressions of ‘clothing’ are more likely to sound natural to participants, enabling me possibly to decrease the high standard deviations for both sentence types. The basic presumption remains the same: I assume significant differences between the test and the control sentences, but also differences between the tested languages depending on their marking of the mass-count distinction. I also still wanted to test the question of whether OMNs – if they are at least minimally acceptable in this kind of construction – are interpreted rather as an individual entity (‘piece of clothing’) or as a bounded collection (‘outfit’). To now be able to cross-check this assumption, I included in this study also sentences with ‘neutral’ verbs that do not necessarily apply to one or the other reading. These are the sentences for French:

(41) Neutral verbs

- a. De plus en plus de gens font du shopping dans les magasins en ligne pour <acheter *une* fringue_{OMN}/*un* vêtement_{SON}>, un livre ou d'autres objets du quotidien.
'More and more people are going online to <buy a clothing/a piece of clothing>, a book or other everyday items.'
- b. Mon père est très grand, à chaque fois qu'il <essaie *une* fringue_{OMN}/*un* vêtement_{SON}> dans un magasin, c'est trop court quelque part.
'My father is very tall, every time he <tries on a clothing/a piece of clothing> in a shop, it's too short somewhere.'
- c. Grand-mère Bouchard aime beaucoup le travail manuel, en ce moment elle <coud *une* fringue_{OMN}/*une* tenue_{CCN}> pour son petit-fils.
'Grandma Bouchard loves to do handicraft, she is currently <sewing a clothing/an outfit> for her grandson.'

(42) Distributive verbs

- a. Beaucoup de gens économisent de l'argent en réparant des choses cassées, mais tout le monde ne sait pas <rapiécer *une* fringue_{OMN}/*un* vêtement_{SON}> ou réparer une chaise.
'Many people save money by repairing broken things, but not everyone knows how to <patch a clothing/a piece of clothing> or to fix a chair.'
- b. J'ai passé toute ma journée à faire le ménage, du coup il ne me reste qu'à <repasser *une* fringue_{OMN}/*une* chemise_{SON}>.
'I spent the whole day doing housework, but now I just have to <iron a clothing/a shirt>.'
- c. Nicolas est fasciné par l'intelligence des robots modernes, bien qu'ils aient encore besoin de beaucoup de temps pour des tâches quotidiennes comme <plier *une* fringue_{OMN}/*une* chemise_{SON}>.
'Nicolas is fascinated by the intelligence of modern robots, although they still need a lot of time for daily tasks such as <folding a clothing/a shirt>.'

(43) Collective verbs

- a. Marie peut toujours compter sur sa sœur aînée. Aujourd'hui, elle l'aide à <combinaison *une* fringue_{OMN}/*une* tenue_{CCN}> pour son entretien d'embauche.
'Marie can always count on her older sister. Today, she helps her to <combine a clothing/an outfit> for her job interview.'
- b. Julie fait très attention à sa tenue vestimentaire, elle sait que le plus petit accessoire peut <perfectionner *une* fringue_{OMN}/*un* ensemble_{CCN}>.

‘Julie takes great care of her outer appearance, she knows that the smallest accessoire can <perfect a clothing/an outfit>.’

- c. Le magazine “Cosmopolitan” conseille ceci: les bonnes chaussures sont là pour <compléter *une* fringue_{OMN}/*une* tenue_{CCN}>, ne dépendez donc pas d’une seule paire.

‘The magazine “Cosmopolitan” advises this: the correct shoes are there to <complement a clothing/an outfit>, so don’t depend on just one pair.’

As can be seen, the sentences in (42b) and (43b) are the same as those in the main study. I included them also in this follow-up study after discussing them again with the supporting native speakers of French (in this case Benjamin Massot of the University of Tübingen), who did not see anything weird in them. This 3x3 design of French could not be maintained completely in all four languages under study, but the general distribution is the same. Figure 4.3 shows the overall results for the control and the test condition in all four tested languages, where the figures in the columns display the number of evaluations considered.

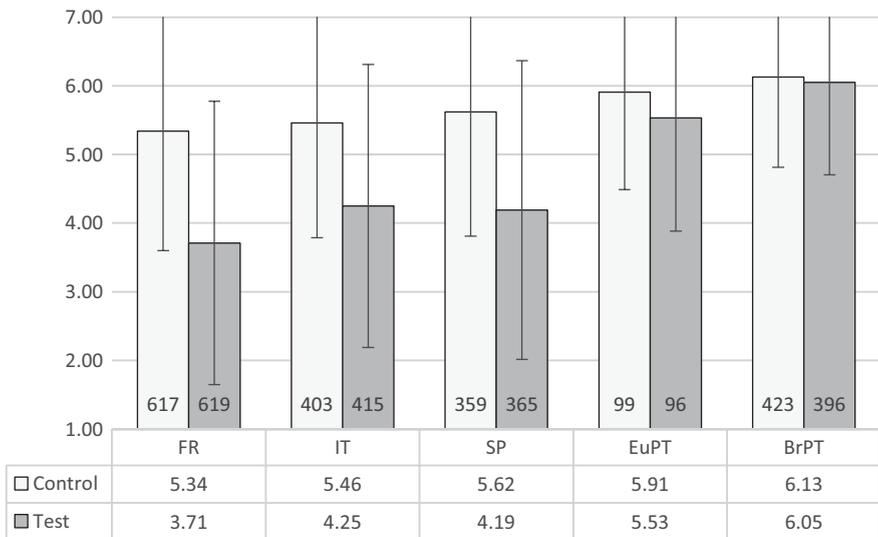


Figure 4.3: Mean acceptability ratings of the construction <indef. art. + OMN> in the four tested Romance languages in the follow-up study.

As shown in Figure 4.3, there is the tendency of Pt. *roupa* ‘clothing’ to be as acceptable as the control condition, while the French, Italian and Spanish OMNs are less acceptable than the singular object noun. To statistically examine this

first observation, the test and the control condition were compared with an independent two-sample t-test in each language. The results reveal that there is a significant difference between the two conditions in French ($t_{FR}(1,200.42) = 15.05$, $p = .000$, $d = .856$), Italian ($t_{IT}(791.38) = 9.23$, $p = .000$, $d = .643$) and Spanish ($t_{SP}(702.73) = 9.65$, $p = .000$, $d = .716$), but not in the two diatopic varieties of Portuguese ($t_{EuPT}(187.15) = 1.71$, $p = .089$; $t_{BrPT}(817) = .91$, $p = .365$) – confirming the first impression of the data. Considering the effect sizes, the difference between the test and the control condition is clearest in the case of French and smallest in the case of Italian, with Spanish occupying an intermediate position. This means that there is indeed a correlation between the marking of the mass-count distinction and the compatibility of OMNs with determiners marking boundedness – but not exactly as was assumed in Table 1.2. According to the nominal systems, one would not expect clear differences between Italian and French but only between Italian and Spanish, as well as between European and Brazilian Portuguese. The overall results thus only confirm the very general tendency of French OMNs to be generally less acceptable than Brazilian Portuguese OMNs, but the continuum between these two poles could not be confirmed. One can therefore conclude that the overall tendencies as assumed in the second semantic-syntactic hypothesis are confirmed, but aspects other than the mere marking of the mass-count distinction may come into play here. One influencing aspect may be the verb type used. As was mentioned previously, it seems to be the case that the verb type influences the range of interpretation and, following from this, also the degree of acceptability of a construction like ‘a clothing’. Taking into consideration only the sentences containing a neutral verb, in other words neither a clear distributive nor a collective reading of the OMN in question, the results are only slightly different (cf. Figure 4.4 below).

Again, an independent two-sample t-test was conducted for each language. The results reveal that there is a significant difference between the test and the control condition in French ($t_{FR}(364.61) = 12.67$, $p = .000$, $d = 1.242$), Italian ($t_{IT}(363.61) = 3.82$, $p = .000$, $d = .399$), Spanish ($t_{SP}(384.36) = 8.75$, $p = .000$, $d = .854$) and also in European Portuguese ($t_{EuPT}(77.63) = 2.50$, $p = .014$, $d = .505$), but exclusively not in the case of Brazilian Portuguese ($t_{BrPT}(271.72) = 1.38$, $p = .169$). Taking into consideration the effect sizes of the significant differences, French again shows the biggest differences between the test and the control condition. Italian represents the other pole of the continuum showing the smallest differences, next to Brazilian Portuguese which does not display any significant differences. Spanish and European Portuguese both lie between the two poles, tending to French and Italian respectively. In light of these results, a continuum of acceptability for the construction ‘a clothing’ would be French < Spanish < European Portuguese < Italian < Brazilian Portuguese. The tendencies of all verb types taken

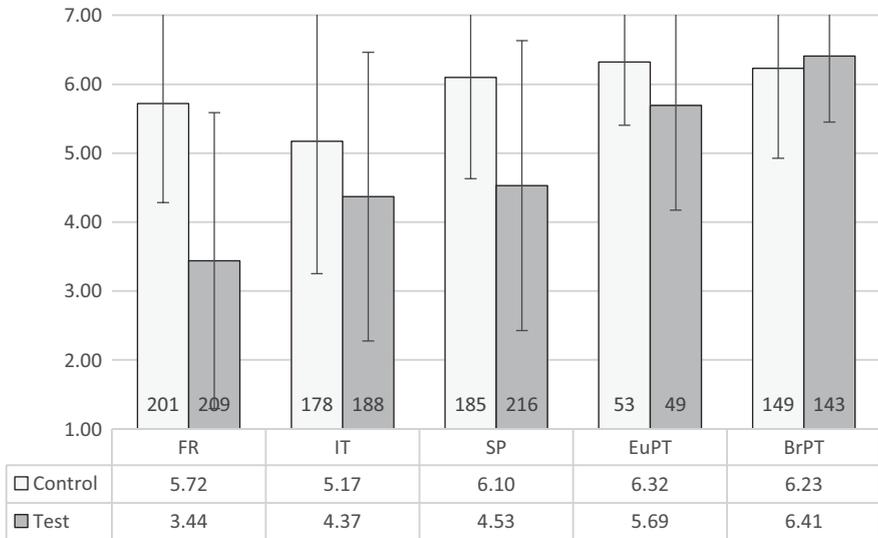


Figure 4.4: Mean acceptability ratings of the construction <neutral verb + indef. art. + OMN> in the four tested Romance languages in the follow-up study.

together could thus be confirmed, but taking into consideration only the neutral verb type, the assumed language continuum corresponds to a higher degree to the assumed continuum of the marking of the mass-count distinction. An exception is represented by It. *abbigliamento* ‘clothing’ which seems to be combinable with the indefinite article, at least to a certain extent. This latter aspect finally leads to a consideration of the other two verb types tested. Given the fact that although there are significant differences between the test and the control condition in nearly all languages studied here, most ratings were not at the impossible end of the scale, but somewhere in the intermediate space. If the tested OMNs can be, if not perfectly then at least relatively acceptable in combination with the indefinite article, are they then interpreted as referring to individual pieces of clothing or rather to a bounded set? To test this assumption, an independent two-sample t-test was conducted for each language and remaining verb type, i.e. distributive vs. collective verbs. For the distributive verbs, the tests reveal significant differences between the test and the control condition for French ($t_{FR}(409.57) = 4.96$, $p = .000$, $d = .485$), Italian ($t_{IT}(236.27) = 10.28$, $p = .000$, $d = 1.261$), Spanish ($t_{SP}(162) = 2.87$, $p = .005$, $d = .451$), but not for European ($t_{EuPT}(40) = .85$, $p = .401$), nor for Brazilian Portuguese ($t_{BrPT}(290) = 1.20$, $p = .232$). For the collective verb condition, the tests reveal significant differences between the test and the control condition for French ($t_{FR}(405) = 9.27$, $p = .000$, $d = .919$), Italian ($t_{IT}(177.22) = 2.87$, $p = .005$, $d = .421$), Spanish ($t_{SP}(157) = 5.81$, $p = .005$, $d = .921$) and again not for

European ($t_{\text{EuPT}}(49) = .55, p = .584$), nor for Brazilian Portuguese ($t_{\text{BrPT}}(233) = 1.59, p = .113$). These comparisons may be interpreted as follows: in both diatopic varieties of Portuguese, *roupa* ‘clothing’ is as acceptable as a SON like *peça de roupa* ‘piece of clothing’ in combination with the indefinite article. It is furthermore also as acceptable as a CCN like *conjunto* ‘outfit’ in combination with a collective verb and the indefinite article. This leads to the conclusion that Pt. *roupa* is not an OMN but a flexible noun, able to express an indefinite, unbounded quantity of clothing, a definite, contextually bounded set as well as an individual item of clothing. In contrast, the lower effect size with the distributive verbs and the higher effect size with the collective verbs suggest that Fr. *fringues* ‘clothes’ as well as Sp. *ropa* ‘clothing’ seem rather to be interpreted as object nouns and not as bounded sets. It. *abbigliamento* ‘clothing’ finally is interpreted rather as a CCN and not as a SON, given the smaller effect size in the case of the collective verbs. This may also be the explanation for the generally higher degree of acceptability of *abbigliamento* with the indefinite article: in contrast to e.g. Sp. *ropa*, It. *abbigliamento* seems to oscillate somewhere between an OMN and a CCN. These impressions are confirmed by a two-way ANOVA with the factors *verb type* (taking into consideration only the distributive and the collective verbs) and test condition: depending on the test condition, the distributive verb condition was rated significantly better than the collective verb condition in French and Spanish (FR: $M_{\text{distr}} = 4.90; M_{\text{coll}} = 4.09; F(1, 822) = 40.16, p = .000, \eta_p^2 = .047$; SP: $M_{\text{distr}} = 5.11; M_{\text{coll}} = 3.69; F(1, 319) = 51.48, p = .000, \eta_p^2 = .139$). In Italian, the collective verb condition was rated significantly better than the distributive verb condition ($M_{\text{distr}} = 4.62; M_{\text{coll}} = 5.34; F(1, 448) = 21.83, p = .000, \eta_p^2 = .046$). The interaction between the two factors is, however, always only small in French and Spanish (FR: $F(1, 822) = 11.87, p = .001, \eta_p^2 = .014$; SP: $F(1, 319) = 6.48, p = .011, \eta_p^2 = .020$) and intermediate in Italian ($F(1, 448) = 20.36, p = .000, \eta_p^2 = .043$). This means, that the differences between the two verb types depending on the test condition are smaller in the case of French and Spanish and bigger in the case of Italian.

Summarising this first test 1a several aspects could be confirmed, while others have yet to be examined in more detail. The analysis of the combination of various collection nouns in the Romance languages was able to confirm the tendency of an assumed continuum of acceptability of these nouns with the indefinite article. These results lead to the conclusion that the marking of the mass-count distinction indeed correlates with the flexibility of the mass noun properties of an OMN. One can even go so far as to say that Brazilian Portuguese, lacking such an overt distinction, does not make use of OMNs but only of flexible nouns. This would empirically confirm the hypothesis of Doetjes (2011, 2564) that “collective mass nouns, which have a count meaning but the morphology of a mass noun, typically occur in languages with an obligatory system of sin-

gular/plural marking, as the lack of number marking distinguishes them from other nouns with count interpretations.” This impression is confirmed by data on Réunion Creole, where we conducted a parallel acceptability judgement study using the test sentences of the main study. Like Brazilian Portuguese, Réunion Creole does not mark the inflectional plural obligatorily, bare nouns are possible in many syntactic contexts and it has no determiner overtly marking mass syntax (cf. Staudacher-Valliamée 2004; Chaudenson 2007; Bollée 2013). The tested translation equivalents for ‘clothing’ *linz* and ‘cattle’ *zanimo* were rated to be as good as the control variant, implying that the nominal system does indeed influence the aspectual constitution of a noun (cf. Kleineberg 2021). Future research should broaden the testing to more nouns in more semantic domains, since the acceptability of the single nouns tested in this present framework may have been influenced by factors other than their syntactic properties, e.g. the colloquial connotation of Fr. *fringues* ‘clothes’. For the semantic domain of CLOTHING, the second semantic-syntactic hypothesis could nevertheless be confirmed.

Test 1b: The second test to examine the characteristics of external boundedness of OMNs is the possibility of the inflectional plural as well as compatibility with more or less distributive quantifiers. In this respect, the continuum of marking of the mass-count distinction in Romance languages not only leads to the assumption of a varying combinability with the indefinite article, but also of a varying acceptability of the inflectional plural – both features refer to the mass noun properties of the collection noun in question. For this reason, just as in the case of test 1a, the acceptability of the OMNs under study with the inflectional plural will be examined in what follows for Spanish, Italian and Portuguese in all three semantic domains. French has to be excluded from this test series since two of the three tested OMNs already have a fossilised plural. The two POMNs in French, *gens* ‘people’ and *fringues* ‘clothes’ will, however, be considered with respect to their combinability with various distributive quantifiers. This approach will address the first semantic-syntactic research question (RQ_{ss}1) and the degrees of countability of POMNs as well as the second semantic-syntactic research question (RQ_{ss}2) and the cross-linguistic variation of mass noun properties of OMNs. To examine these two aspects of analysis, test sentences which again do not allow any kind of sortal reading were constructed. Just as in the case of the indefinite article tests, a sortal reading may also well be possible with an inflectional plural (cf. Lauwers 2016). The first part of the tests examines the acceptability of the inflectional plural with SOMNs in Italian, Spanish and Portuguese. The test construction here is <non-distributive quantifier + SOMN_{PL}>, since the test should concentrate only on the inflection and not on any kind of determiner, and the control condition is repre-

sented here by a plural object noun. The second part of the tests then focuses on the two POMNs in French and their acceptability in the two constructions <small round number + POMN> and <distributive quantifier + POMN>, viz. highly distributive numerals vs. a less distributive *fuzzy denominator* (cf. Allan 1980). These two test conditions are compared to the control condition with the POMN and a non-distributive quantifier. I chose test sentences where the two options are plausible and only exchanged the quantifier, and the test sentences were constructed in a parallel way for all languages under study. The following examples illustrate the plural inflection test sentences exemplarily for Spanish and the distributive quantifier sentences for French (PON = plural object noun):

- (44) Inflectional plural – cross-linguistic comparison of SOMNs
- a. CLOTHING: Es absolutamente necesario que ordene mi armario. ¡Tengo <*muchas* ropas_{SOMN-PL}/prendas_{PON}> que nunca me pongo!
‘I really have to tidy out my wardrobe. I have <many clothing_{PL}/pieces of clothing> which I never have worn!’
 - b. CATTLE: Al pobre veterinario Díaz le queda mucho por hacer, aún tiene que examinar <*muchos* ganados_{SOMN-PL}/animales_{PON}>.
‘The poor veterinarian Díaz has much to do, he still has to examine <many cattle_{PL}/animals>.’
 - c. PEOPLE: Hoy en el autobús nadie estaba dispuesto a cederle su asiento a una mujer embarazada, aunque todavía había espacio para <*muchas* gentes_{SOMN-PL}/personas_{PON}> en el fondo.
‘Today on the bus nobody wanted to offer their seat to a pregnant woman although there was still enough space for <many people_{PL}/persons> in the back.’
- (45) Distributive quantifiers – French POMNs
- a. J’ai vraiment besoin de faire le vide dans mon armoire. Je vois déjà <*cinq* fringues_{POMN}/vêtements_{PON}>//Il y a <*beaucoup de/plusieurs* fringues_{POMN}/vêtements_{PON}> que je n’ai jamais porté(e)s !
‘I really have to tidy out my wardrobe. I see already <five clothes/pieces of clothing>//there are <many/various clothes/pieces of clothing> which I have never worn!’
 - b. Aujourd’hui dans le bus, personne n’était prêt à laisser sa place à une femme enceinte, alors qu’il y avait encore de la place pour <*beaucoup de/cinq/plusieurs* gens_{POMN}/personnes_{PON}> au fond.
‘Today on the bus nobody wanted to offer their seat to a pregnant woman although there was still enough space for <many/five/several people/persons> in the back.’

As can be seen in examples (44) and (45), the sentences for both test domains as well as for the control and the test conditions are exactly the same; the only things that have changed are the respective nouns and quantifiers. In the case of Brazilian Portuguese, I did not include the semantic domain of PEOPLE, since *gente* was evaluated very low in all test contexts, suggesting that the already high degree of pronominalisation hinders its interpretation as a full noun (cf. chap. 4.1). In the case of the construction <small round number + (POMN/PON)_{inanimate}> in French, the test sentence was altered slightly to make it sound more natural. I will first discuss the results for the inflectional plural of SOMNs and its cross-linguistic variation and will then discuss the two POMNs in French and their combinability with more or less distributive quantifiers. The results for the evaluations of the various SOMNs with inflectional plural and non-distributive quantifier are displayed in the following Figure 4.5, the figures in the columns again represent the number of evaluations per condition.

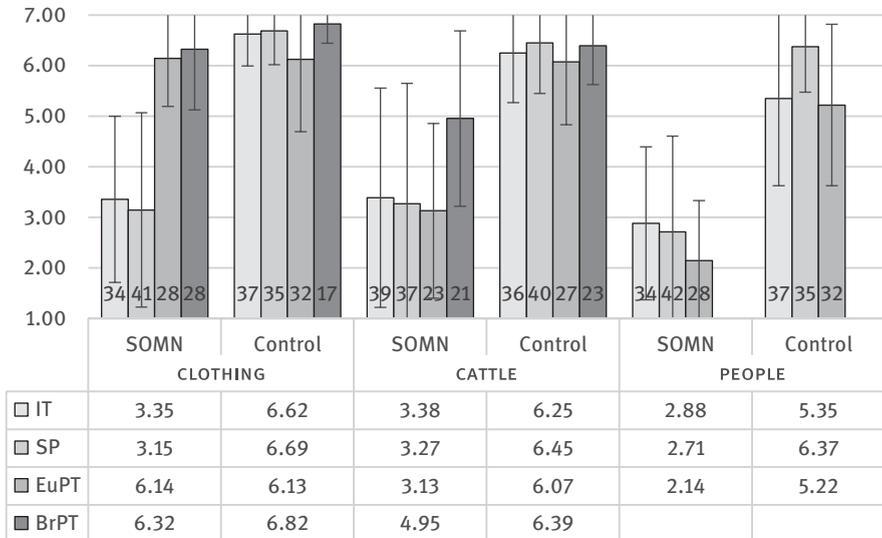


Figure 4.5: Mean acceptability ratings of the construction <non-distributive quantifier + SOMN_{pl}> in Italian, Spanish and Portuguese.

A first observation of this graph leads to two preliminary impressions: the SOMNs are all rated relatively low in comparison to the overall good control condition. At first sight, the systematic cross-linguistic variation determined in the case of test 1a cannot be confirmed. Note that Pt. *roupa* ‘clothing’, which was rated as acceptable with the indefinite article as the control condition, again shows rela-

tively high mean ratings in both European and Brazilian Portuguese. To examine these impressions in more detail, an independent two-sample t-test was conducted for each of the three semantic domains and the four languages. In accordance with the hypothesis of systematic cross-linguistic variation with respect to the mass noun properties of the tested SOMNs, I assume significant differences between the test and the control condition in Italian, no significant differences between the two conditions in Brazilian Portuguese, and that Spanish and European Portuguese may lie somewhere in between. In the domain of CLOTHING, there is a significant difference between the test and the control condition in Italian and Spanish ($t_{IT}(41.81) = 10.73, p = .000, d = 2.631$; $t_{SP}(50.94) = 10.91, p = .000, d = 2.358$), but not in the case of European and Brazilian Portuguese ($t_{EuPT}(58) = .055, p = .956$; $t_{BrPT}(35.29) = 2.01, p = .052$). In the domain of CATTLE, there is a significant difference between the test and the control condition in all languages being studied ($t_{IT}(53.93) = 7.37, p = .000, d = 1.658$; $t_{SP}(47.51) = 7.44, p = .000, d = 1.744$; $t_{EuPT}(48) = 6.84, p = .000, d = 1.940$; $t_{BrPT}(26.97) = 3.43, p = .002, d = 1.067$). This also holds for the domain of PEOPLE: there is a significant difference between the test and the control condition in Italian, Spanish and European Portuguese ($t_{IT}(69) = 6.30, p = .000, d = 1.496$; $t_{SP}(60.83) = 10.97, p = .000, d = 2.369$; $t_{EuPT}(58) = 8.23, p = .000, d = 2.130$). Summarising these results, the tested SOMNs are significantly less acceptable with the inflectional plural than the object noun control variant. This holds for all semantic domains and languages tested, irrespective of their nominal system and supposed flexibility of mass noun properties of a SOMN. An exception is represented by Port. *roupa*, which is broadly acceptable with the inflectional plural, confirming its flexible status as already assumed in test 1a. These results refute the cross-linguistic continuum as assumed in the second semantic-syntactic hypothesis (HP_{ss2}). There are two issues with these test results: firstly, for various methodological reasons, the two extreme poles of the assumed continuum of the marking of the mass-count distinction could not be included in this test series, or at least only to a certain extent. This concerns mainly French, where the primarily tested POMNs already imply plurality and therefore had thus to be excluded, and Brazilian Portuguese where the human domain had to be excluded due to the high degree of pronominalisation of *gente*. As shown in Table 1.2, it is exactly French and Brazilian Portuguese that represent the two extreme poles of marking of the mass-count distinction. This makes the hypothesis of a systematic cross-linguistic variation of the mass noun properties of SOMNs more difficult to examine. Secondly, the fact that there are no differences between European Portuguese and Italian and Spanish in the domain of CATTLE and PEOPLE – in contrast to the test results of test 1a – may lead to two possible conclusions: either the inflectional plural is not as susceptible to the mass-count distinction continuum as the indefinite article, or noun-specific

idiosyncrasies are of greater importance for their acceptability in the tested constructions than the different nominal systems. An argument in favour of the latter hypothesis is the better evaluation of Pt. *gado* in Brazil, but not in Europe – in the former variety, cattle are presumably much more present in people’s everyday lives. This hypothesis is however difficult, even impossible, to examine. This first test domain of test 1b thus unfortunately leads to inconclusive results, but it once again stresses the importance of a global perspective on OMNs. As has already been stated several times, most research on English OMNs has been conducted with *furniture* as an exemplar; this present study, however, clearly shows important differences between various nouns of this category, inter- and intra-linguistically. Taking together the results of test 1a and this first part of test 1b, one might summarise that the assumed continuum of the flexibility of mass noun properties is limited to certain exemplars of the category of SOMNs. It may be that it is limited to the semantic domain of artefacts or even only to the domain of CLOTHING, and future research should concentrate on studies that take into consideration more exemplars from various semantic domains to examine this more thoroughly.

The second part of test 1b addresses the combinability of the French POMNs *gens* ‘people’ and *fringues* ‘clothes’ with more or less distributive quantifiers. The results of the evaluations of the test sentences in (45) are displayed in Figure 4.6:

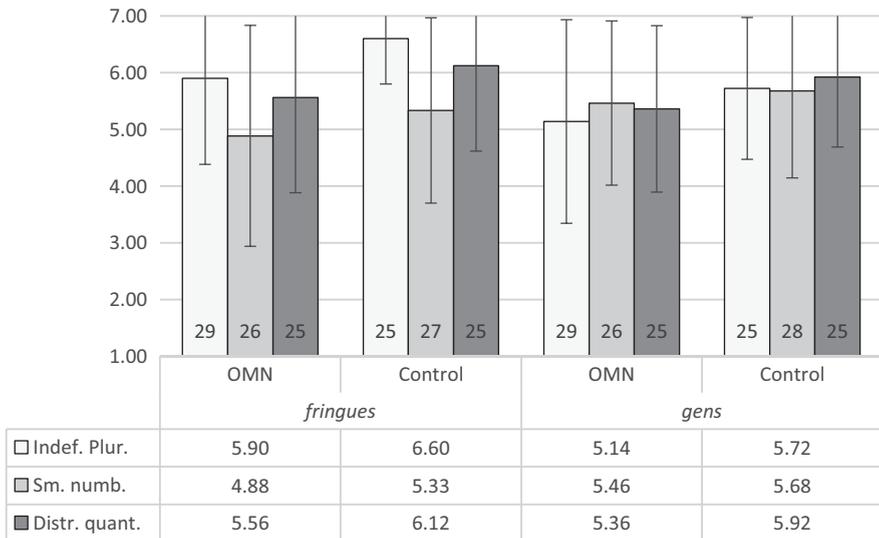


Figure 4.6: Mean acceptability ratings of French POMNs in combination with small round numbers vs. *plusieurs* ‘various’.

Figure 4.6 shows for each of the two tested nouns the results of the control condition and the two test conditions, respectively for the sentences with the cardinal number and the distributive quantifier *plusieurs* ‘various’, the numbers at the bottom of each column again representing the number of evaluations for each sentence. The mean acceptability ratings for the two POMNs in combination with the two kinds of quantifier as displayed in the graph already indicate that there is no notable difference between the test and the control conditions. To test the statistical significance of this impression, a two-way ANOVA with the factors *condition* and *type of quantifier* was conducted for each of the two POMNs. This analysis reveals that there is a weak main effect on *condition* in the case of *fringues* ($F(1, 151) = 5.06, p = .026, \eta_p^2 = .032$), an intermediate main effect on the type of quantifier ($F(2, 151) = 7.04, p = .001, \eta_p^2 = .085$), but no significant interaction between the two main factors ($F(2, 151) = .086, p = .917$). This means that there are indeed differences between the three tested quantifier conditions, but that the differences between the POMNs and the plural object nouns are equal in every condition. It follows from this that *fringues* is equally acceptable with a non-distributive quantifier like *beaucoup* ‘much/many’, a highly distributive quantifier like a small round number as well as other distributive quantifiers like Fr. *plusieurs* ‘various’. The same holds for *gens*. Here, the results are even clearer: there is no main effect on condition ($F(1, 152) = 3.56, p = .061$), and nor is there a main effect on the type of quantifier ($F(2, 152) = .27, p = .766$), nor is there any interaction between the two main factors ($F(2, 152) = .25, p = .782$). These results thus support the argument of Lauwers (2014) that these POMNs are indeed acceptable with both a small round number and a distributive quantifier of the kind ‘several/various’. Given the good results of *gens*, I ascribe the slightly smaller acceptability of *fringues* to its colloquial connotation. Interestingly, the results also refute the (prescriptive) position of Grevisse/Goosse (2016, §510) that *gens* is not acceptable with cardinal numbers.

Summarising the results of tests 1a and 1b with respect to the second semantic-syntactic research question (RQ_{ss2}), one can now state the following: the tests were designed with the hypothesis that restrictions on the countability of Romance OMNs – correlating with their external unboundedness – depend on the nominal system of the language in question. In specific terms, it has been assumed that French and Italian, because they mark mass nouns overtly with the partitive article, do not allow any kind of syntactic flexibility of OMNs, while especially Brazilian Portuguese, not marking mass nouns syntactically, was assumed to show few, if any, restrictions. To test this hypothesis, various OMNs in the semantic domain of CLOTHING were combined with the indefinite article in various contexts. The results confirm the assumed continuum for these nouns, but only to a certain extent. Whereas the assumed low acceptability for Fr. *une fringue* ‘a clothing’ and the perfect acceptability of BrPt. *uma roupa* ‘a clothing’

could be confirmed, It. *abbigliamento* was surprisingly well accepted in this kind of construction. It was also tested how far the verb type used in these constructions influences the interpretation of ‘a clothing’. It could be shown that Fr. *fringues* and Sp. *ropa* were interpreted (or perhaps coerced) rather as a singular object noun and It. *abbigliamento* as a CCN. The cross-linguistic continuum could not, however, be confirmed in the second test, which examined the possibility of the inflectional plural with various SOMNs. In all languages and all semantic domains tested, the test condition was rated significantly lower than the control condition. An exception was represented by Pt. *roupa*, confirming its status as a flexible noun. As a last test domain, the hypothesis was examined regarding the degree to which POMNs are compatible with more or less distributive quantifiers, addressing the first semantic-syntactic research question, and possible correlations between the external boundedness and the accessibility of the internal plurality of a collection. The results confirm the assumption of Lauwers (2014) that lexical plurals are perfectly acceptable with distributive quantifiers. Following on from this, one may conclude that the massifying effect of a POMN is less strong than that of a SOMN, and the collection may well be dissolved to count the constituting entities. Taking together all results of tests 1a and 1b, the hypothesis of countability preferences as assumed by Allan (1980) and Grimm (2018) can be confirmed. Whereas some nouns are flexible with respect to their aspectual status (e.g. Pt. *roupa*), some nouns like It. *abbigliamento* may be acceptable with the indefinite article in the singular, but not with the inflectional plural. This leads to the conclusion that the assumed countability continuum may indeed be dependent on the specific nominal system of a language, but also on other factors such as the fossilised number (SOMN vs. POMN). Factors such as usage frequency and the semantic domain may also come into play, but this should be examined in more detail in future research.

4.2.2 Testing the internal plurality of count collective nouns and object mass nouns

Having elaborated on the external unboundedness of Romance OMNs, this second empirical chapter will address issues relating to the internal plurality of collection nouns. As indicated above, I tested, on the one hand, the accessibility of the constituting referents of a collection by combining in my tests the noun in question with stubbornly distributive predicates as well as with highly distributive constructions like ‘one after the other’. On the other hand, I tested the hierarchical relations between the noun denoting the collection and its constituting entities, *viz.* whether it is a hyponymic or a meronymic relation (or both or neither

of the two). I did not test collections in constructions *ad sensum* because, as has been described in chap. 3.1.2, there are too many factors influencing the possibility of these constructions (e.g. the distance between subject and predicate). Constructions *ad sensum* may thus be well examined in corpus analyses, but they are less suited for neatly designed acceptability judgement tests.

Test 2a: In accordance with the first semantic-syntactic hypothesis elaborated above, I assume with respect to compatibility with stubbornly distributive predicates a good acceptability of this construction for OMNs, expected to be as good as the control condition, and a significantly lower degree of acceptability for CCNs. These assumptions are based on the categorisation of OMNs as (mostly) unbounded collections and CCNs as bounded ones. Both kinds of collection are constituted of a plurality of entities, but only in the former case are they accessible by means of combination with the respective collection noun and a stubbornly distributive predicate. Since this compatibility of OMNs with stubbornly distributive predicates was equally assumed for English and Portuguese OMNs, there is nothing that would predict differences between the Romance languages, as was determined by compatibility with the indefinite article. For this reason, I concentrate in the following on the results for French, the language focused on in this present work. In addition, I also present the results of the Italian testing since there could indeed be differences between the (mostly) tested POMNs in French and the exclusively tested SOMNs in Italian, both being comparable on the level of their nominal system. In both cases, the OMNs of Table 4.3 were contrasted with a CCN as a second test variant and a plural object noun (PON) as the control variant. In the French study, I could also include in this testing the flexible noun Fr. *vêtement*, which in the singular may bear the interpretation of either ‘piece of clothing’ or ‘clothing’, depending on the syntactic context. The test was carried out for all three semantic domains:

- (46) a. FR: Quand je suis arrivé à Berlin, il faisait si froid que j’ai enfilé immédiatement <des fringues_{POMN}/du vêtement_{SOMN}/un ensemble_{CCN}/des vêtements_{PON} *longues/long(s)*>.
 IT: Quando sono arrivato a Lanzarote, faceva così caldo che mi sono messo subito <un abbigliamento_{SOMN}/un outfit_{CCN}/dei vestiti_{PON} *corto/i*>.
 ‘When I arrived in Berlin/Lanzarote, it was so cold/hot that I immediately <put on long/short clothing/a long/short outfit/long/short pieces of clothing>.’
- b. FR: L’agriculteur Dubois se fait de la bile à cause <du bétail_{SOMN}/du troupeau_{CCN}/des bêtes_{PON} *maigre(s)*>, les conditions de cette année étaient vraiment mauvaises.

IT: L'agricoltore Martini si preoccupa per <il bestiame_{SOMN}/il gregge_{CCN}/le bestie_{PON} *magro/e*>, quest'anno le condizioni erano davvero pessime. 'The farmer Dubois/Martini is worried about the <lean cattle/herd/animals>, this year's conditions were really bad.'

- c. FR: Isabelle envie <les gens_{POMN}/femmes_{PON} *minces*> qui peuvent manger beaucoup tout en ne pas grossissant.//Isabelle montre des photos de son séjours aux États-Unis à Lisa, elle envie un peu <l'équipe_{CCN} *mince*> de cheerleaders.

IT: Isabella invidia <la gente_{SOMN}/le donne_{PON} *snella/e*> che può/possono mangiare molto senza ingrassare.//Isabella mostra a Lisa delle foto del suo soggiorno negli Stati Uniti, invidia un po' <la squadra_{CCN} *snella*> di cheerleader.

'Isabel is envious of <the slim people/women> who can eat a lot without getting fat.//Isabel shows some photos of her visit to the USA to Lisa who is a little envious of <the slim cheerleading team>.'

The test sentence (46c) containing the CCN Fr. *équipe*/It. *squadra* 'team' was slightly altered to create a logical context for it. Figure 4.7 summarises the mean acceptability ratings as well as the number of rated sentences, the standard deviations and the number of ratings for these constructions.

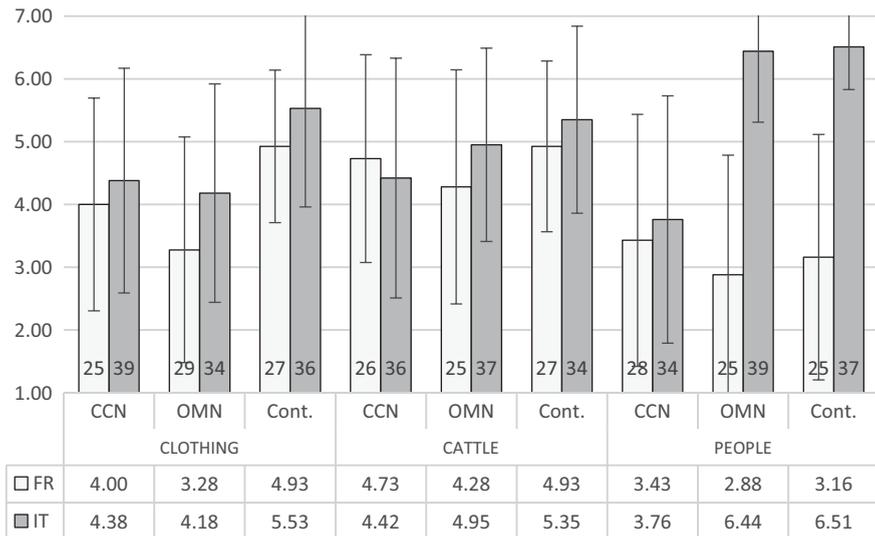


Figure 4.7: Mean acceptability rating of French and Italian collection nouns in combination with a stubbornly distributive predicate.

As shown in Figure 4.7, the results for this testing domain are very heterogeneous. In the domain of CLOTHING, the mean ratings of all three sentence types are only mediocre, and even the control condition was not rated well. The same holds for the domain of CATTLE and PEOPLE in French. In the case of Italian, the results are better, displaying a smaller standard deviation ($M_\sigma = 1.54$ vs. $M_\sigma = 1.72$ in French) and good control sentences. Especially the generally mediocre control condition in French, however, leads to the fact that these results are not well suited for further (statistical) analysis. It may be that idiosyncratic features come into play here. For instance, the context testing the construction *long clothing* would have been more natural with other adjectives like *warm*. In addition, the context examining the construction *slim people* implies perhaps mostly women. For Fr. *gens* not being specified for gender, this implication could have been weird for some participants (cf. also chap. 8.3.2 for corpus data on this). This, however, does not explain the good ratings of *gente snella* in Italian. In addition, we were not sufficiently aware of the fact that It. *abbigliamento* was presented with a partitive article, just as the French translation equivalents. Because of these inconsistencies, we thus conducted a follow-up testing. The semantic domain of PEOPLE was excluded in this study for the reasons mentioned. The domain of CATTLE was also excluded because the CCNs *troupeau maigre* and *gregge magro* could possibly have been reinterpreted as ‘small herd’ (and not as ‘herd being constituted of lean cattle’). Instead, I examined the constructions *long clothing* (in a more idiosyncratic context), *rectangular crockery* and *rounded furniture*. These OMN-variants were again contrasted with a plural object noun (PON) and a CCN. In the follow-up study it was taken care that all nouns were presented with the definite article. The sentences tested are cited in (47):

- (47) a. FR: En plus <du vêtement_{SOMN}/des fringues_{POMN}/de l'ensemble_{CCN}/des vêtements_{PON} *long(ues)*>, il faudrait aussi que vos enfants mettent un antimoustique pour se protéger encore mieux des piqûres d'insectes.
 IT: Oltre <all'abbigliamento_{SOMN}/all'outfit_{CCN}/ai capi di abbigliamento_{PON} *lungo/hi*> i bambini dovrebbero utilizzare un repellente antizanzare per proteggersi il più possibile dalle punture.
 ‘In addition to <the long clothing/outfit/pieces of clothing>, your children should apply a mosquito repellent to be best protected against bites.’
- b. FR: La semaine dernière, on est allé dans un restaurant très branché, mais <la vaisselle_{SOMN}/le service_{CCN}/les assiettes_{PON} *carré(es)*> et la décoration élégante n'ont pas pu détourner l'attention du mauvais repas.
 IT: La settimana scorsa siamo andati in quel ristorante super chic all'angolo, ma <il vasellame_{SOMN}/le stoviglie_{POMN}/il servizio_{CCN}/i piat-

ti_{PON} *quadrato/e/i*> e l'arredamento elegante non sono riusciti a compensare la cattiva qualità del cibo.

'Last week, we went to a fancy restaurant, but <the rectangular crockery/dishes/service/plates> and the elegant interior design could not distract from the horrible food.'

- c. FR: Marie-Claire a réalisé son rêve en rénovant un ancien phare comme maison de vacances, c'est même elle qui a fabriqué <le mobilier_{SOMN}/l'ameublement_{CCN}/les meubles_{PON} *arrondi/s*> que tu as vus dans son atelier hier.

IT: Finalmente Riccardo ha realizzato il suo sogno e si è comprato un vecchio faro come seconda casa e si è addirittura costruito da solo <della mobilia_{SOMN}/dell'arredamento_{CCN}/dei mobili_{PON} *arrotondata/o/i*>. 'Mary/Ricardo fulfilled her/his dream and renovated an old lighthouse as a holiday home. S/he even fabricated <the round furniture/furnishings/pieces of furniture> which you saw yesterday in her workshop.'

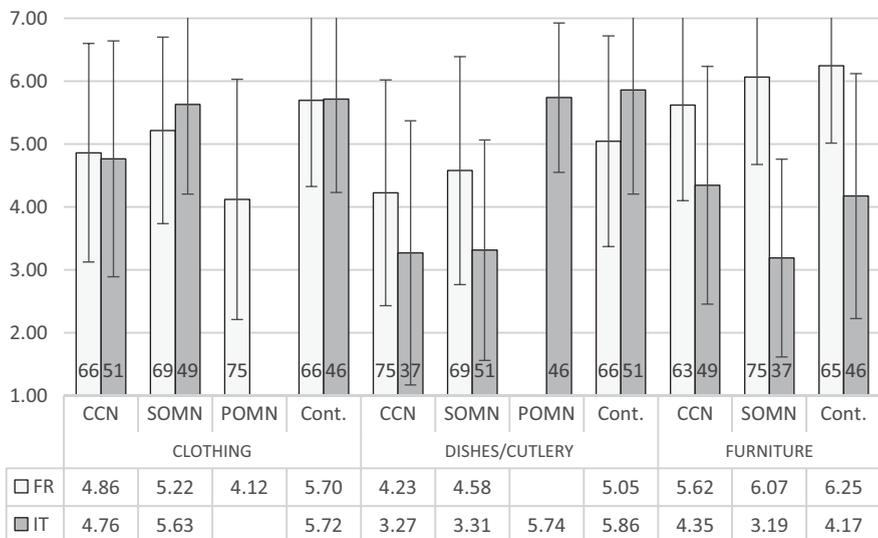


Figure 4.8: Mean acceptability rating of French and Italian collection nouns in combination with a stubbornly distributive predicate in the follow-up-study.

Figure 4.8 shows the mean acceptability ratings of all test sentences in French and Italian. To test the assumption that there is a significant difference between the various nominal types tested with respect to their compatibility with a stubbornly distributive predicate, a one-way ANOVA with the independent factor *con-*

dition was conducted for French and Italian separately, taking into consideration all semantic domains tested. In French, there is a significant difference between the four nominal types: $F(3, 685) = 17.62, p = .000, \eta_p^2 = .072$. A Bonferroni post-hoc test reveals that the CCN condition was rated significantly lower than the control condition ($p = .000$) and the SOMN condition ($p = .045$), and that there is no significant difference between the SOMN- and the control condition ($p = .227$). This confirms the first semantic-syntactic hypothesis that an increasing outer boundedness of a collection correlates with a decreasing accessibility of internal plurality, and CCNs are thus hardly compatible with stubbornly distributive predicates. The French POMN *fringues* ‘clothes’ was furthermore rated significantly lower than the CCN condition ($p = .008$), the SOMN condition ($p = .000$) and the control condition ($p = .000$). At first sight, this is surprising given the fact that a POMN should display a lower degree of boundedness and thus a higher degree of accessibility of internal plurality. The lower ratings of *fringues* are, however, probably due to the colloquial character of the noun. This is supported by the fact that many participants commented about this sentence that it mixes up linguistic styles, the formal style of a teacher/instructor with the informal/colloquial style of *fringues*. In Italian, the one-way ANOVA also showed significant differences between the four conditions: $F(3, 459) = 16.30, p = .000, \eta_p^2 = .096$. Just as in the case of French, the CCN condition was also rated significantly lower than the control condition ($p = .000$). In contrast to French, however, the SOMN condition was also rated significantly lower than the control condition ($p = .000$). This may be because the SOMNs *vasellame* ‘crockery’ and *mobilia* ‘furniture’ were mentioned by participants as being *antico* ‘outdated’ and *inusuale* ‘unusual’. I will thus only consider the good rating of *abbigliamento* ‘clothing’ in this context (cf. Figure 4.8), and interpret the data as confirming the tendency displayed by French that SOMNs are as good with a stubbornly distributive predicate as the control condition. The POMN *stoviglie* ‘dishes’ was equally rated as being good as the control condition, and there were no significant differences between the two conditions ($p = .888$). This also supports the fact that the low evaluation of Fr. *fringues* is due to its colloquial character and not due to its nominal type. The first semantic-syntactic hypothesis can thus be confirmed: the compatibility of a certain collection noun with a stubbornly distributive predicate accessing its internal plurality depends on its nominal type. In this respect, the degree of external boundedness correlates with the accessibility of this internal plurality. CCNs are generally not compatible with this kind of construction since their set profiling hinders access to their constituting entities. In contrast, OMNs – both SOMNs and POMNs – are easily combinable with a stubbornly distributive predicate given their unboundedness. Compared to the results in chap. 4.2.1, which showed that grammatical means of expression of nominal aspectuality are susceptible to the degree

of boundedness also displayed by the morphological form, semantic means of expression like distributive adjectives seem to be less restrictive in this respect.

Test 2b: The second test examining accessibility of the internal plurality of collections tested the construction <collection noun + ‘one after the other’>. Concerning this construction, the state of the art as described in chap. 3.1.2 is not entirely clear. For English, these highly distributive constructions have been said to be ungrammatical in combination with an OMN (cf. Rothstein 2010a, 379–380), whereas for Brazilian Portuguese there indeed seems to exist the possibility of such a construction – although the examples presented are modified nouns like *mobília dessa marca* ‘furniture of this brand’ (cf. Rothstein 2010a, 383–384; Pires de Oliveira/Rothstein 2011a, 2157). For this reason, there is the assumption that this kind of highly distributive construction may indeed be susceptible to cross-linguistic variation, in contrast to stubbornly distributive predicates. The following test domain presents results for all four Romance languages being studied to address the second semantic-syntactic hypothesis elaborated above that Brazilian Portuguese may display a higher acceptability of these constructions than the other Romance languages tested, given its lower degree of grammaticalisation of the mass-count distinction. In French, this test can again also consider the SOMN *vêtement* ‘clothing’ which may be contrasted with the POMN *fringues* ‘clothes’. In accordance with the first semantic-syntactic hypothesis (HP_{ss1}), I contrast the OMN with a CCN and a plural object noun (PON) as control variant. The predictions are broadly similar as in the case of test 2a: since CCNs are bounded entities, I assume them not to be acceptable in these contexts. The internal plurality should not be accessible because of the set profiling of this aspectual type. This unacceptability should be present in all languages, irrespective of their degree of grammaticalisation of the mass-count distinction. For OMNs, I suppose that there may indeed be differences between the Romance languages with a decreasing acceptability from Brazilian Portuguese to French. It is not clear how the POMNs *gens* and *fringues* in French will behave in this testing. Their fossilised plural morphology should facilitate their combination with ‘one after the other’ (similar to the combination with distributive quantifiers); the blurring effect of transnumerality, however, might also impede it. Schnedecker (2012, 146) argues in favour of the latter hypothesis by referring to Wierzbicka (1985, 282–283) who claims that the status of a plurale tantum implies that “the parts may not be truly separate”. Schnedecker deduces from this that *gens* – as a plurale tantum – is not compatible with highly distributive predicates. Although her reasoning cannot be true for a heterogeneous mass noun like *gens* – Wierzbicka refers to pluralia tantum like *groceries* or *leftovers* that may also contain substances (for example soup) – the results may nevertheless be true: the massification effect of the OMN may block

linguistic accessibility of the constituting individuals of *gens*.⁴⁹ In accordance with the third semantic-syntactic hypothesis (HP_{ss3}) and the animacy hierarchy, one may also assume differences between the three semantic domains. Given the assumption that human referents are more salient than non-human and inanimate referents, there should also be a decreasing acceptability of this tested construction from PEOPLE to CATTLE to CLOTHING. Example (48) displays the test and control sentences for French:

- (48) a. La teinturerie Nicolas garantit qu'on contrôlera <les fringues_{POMN}/le vêtement_{SOMN}/la tenue_{CCN}/les vêtements_{PON} l'un(e) après l'autre> avant la livraison.
 'The dry cleaner's Nicolas guarantees that they control <the clothes/the clothing/the outfit/the pieces of clothing one after the other> before the delivery.'
- b. Le vétérinaire Bagot examine <le bétail_{SOMN}/le troupeau_{CCN}/les vaches_{PON} soigneusement l'un(e) après l'autre>.
 'The vet Bagot examines <the cattle/the herd/the cows carefully one after the other>.'
- c. Après l'incident, le commissaire interroge <les gens_{POMN}/l'équipe_{CCN}/les personnes_{PON} l'un(e) après l'autre>.
 'After the incident, the inspector questions <the people/the team/the persons one after the other>.'

In what follows, I will firstly present the results for every single degree of animacy, i.e. inanimate (CLOTHING), animate (CATTLE) and human (PEOPLE), to then come to the comparison of them. Figure 4.9 shows the mean ratings of the construction <collection noun_{inanimate} + 'one after the other'> for all four Romance languages being studied, the figures at the bottom of each column again represent the number of evaluations considered:

⁴⁹ In contrast to this, the construction *gens les uns après les autres*, thus with the pronouns *un* 'one' and *autre* 'other' not in the singular, but in the plural, is acceptable according to Cappeau/Schnedecker (2014, 65). They assume that this construction does not individuate single persons, but creates sub-sets which may then be referred to distributively. In accordance with this, a sentence like *les gens arrivent les uns après les autres* has a meaning like 'the people arrive the ones after the others, i.e. first the students, then the professors and then the parents' and not 'first Sabrina, then Christian, then Thomas'.

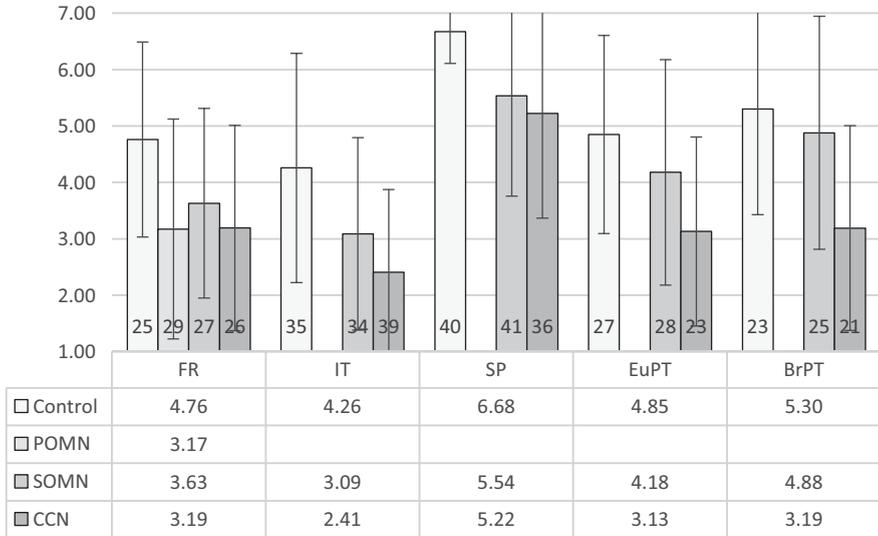


Figure 4.9: Mean acceptability rating of inanimate Romance OMNs and CCNs in combination with highly distributive predicates.

Figure 4.9 shows several kinds of difference between the types of collection noun tested in comparison to the plural object noun control condition. To analyse these differences, a one-way ANOVA with the factor *condition* on each of the tested languages was conducted – in this first step I had not yet included the factor *language*. The ANOVA reveals for every tested language a main effect of *condition* ($F_{FR}(3, 103) = 4.24, p = .007, \eta_p^2 = .110$; *Welch's* $F_{IT}(2, 67.02) = 9.53, p = .000, \eta_p^2 = .221$; *Welch's* $F_{SP}(2, 57.74) = 15.77, p = .000, \eta_p^2 = .353$; $F_{EuPT}(2, 75) = 5.34, p = .007, \eta_p^2 = .125$; $F_{BrPT}(2, 66) = 7.01, p = .002, \eta_p^2 = .175$). Although the ANOVAs are all statistically significant and η_p^2 indicates intermediate to large effect sizes, an important note on the standard deviation is necessary: as shown in Figure 4.9, the variation between the different ratings is very high, and this also applies for the control conditions. This observation also had to be made for several other tests discussed so far. The relatively high standard deviation was present in both the main and the follow-up study, so the relatively large number of test sentences could thus not be responsible for the high degree of variation. Besides the possibility of diverging individual anchoring points on the evaluation scale, a comment of one of the Italian participants sheds light on this issue. He or she wrote: “Aggiungere *una dopo l'altra* è superfluo e la frase non suona più molto naturale” (‘to add *one after the other* is superfluous and the sentence does not sound very natural anymore’). The person did not rate any of the control sentences (but did rate e.g. the OMN con-

dition in the animate domain), but the comment may be considered as valid for all conditions. In other words, although a certain construction may be possible, this does not necessarily mean that it is a natural construction for a native speaker of the language in question. This not only applies for this particular test but also for the other constructions being studied. This high degree of variance in the data slightly decreases the reliability of the results in general, but they still are clear: the post-hoc tests (Bonferroni in the case of the ANOVAs and Games-Howell in the case of the Welch ANOVAs) reveals that there is a statistically significant difference between the control condition and the CCN condition in every language studied here ($p_{FR} = .018$; $p_{IT} = .000$; $p_{SP} = .000$; $p_{EuPT} = .005$; $p_{BrPT} = .002$). This clearly proves the hypothesis formulated above that the set profiling of CCNs hinders accessibility of the internal plurality of the collection. Spanish and Italian additionally seem to pattern with English, since there is also a significant difference between the control condition and the SOMN condition ($p_{IT} = .035$; $p_{SP} = .001$). In contrast, both diatopic varieties of Portuguese do not display any significant differences between these two conditions ($p_{EuPT} = .553$; $p_{BrPT} = 1.000$). This confirms the impression of Rothstein (2010a, 383–384) and of Pires de Oliveira/Rothstein (2011a, 2157) that these highly distributive constructions indeed work well in Portuguese; this is not only true for the Brazilian variety but also for European Portuguese. In light of these results, one may indeed confirm the cross-linguistic variation postulated by Rothstein (2010a) and one could also add an explanation for it: the acceptability of a highly distributive predicate in combination with a SOMN correlates with the degree of grammaticalisation of the mass-count distinction. Three incongruencies are, however, of importance here: except for Brazilian Portuguese, there is no significant difference between the SOMN and the CCN condition in any of the tested languages ($p_{FR} = .1000$; $p_{IT} = .183$; $p_{SP} = .737$; $p_{EuPT} = .124$; $p_{BrPT} = .016$). In light of the assumed continuum of aspectual types as summarised in Figure 3.1, one would expect systematic differences between these two nominal types, given the distinction between a bounded and an unbounded collection. Furthermore, Spanish mainly patterns with European Portuguese with respect to the coding of the mass-count distinction (cf. Table 1.2). Against this background, one would tend to assume differences between the two varieties of Portuguese and not between European Portuguese and Spanish. It could be that the possibility of *aspectual singulars* in European Portuguese, and to a lesser extent in Spanish, as one syntactic feature blurring the mass-count distinction may play a role here. As was shown in the other tests, a one-to-one mapping of the assumed continuum cannot be confirmed. On the other hand, the French data do not fit the picture at all. There is no significant difference between the SOMN *vêtement* ‘clothing’ and the control variant ($p = .172$), but

there is a significant difference between the POMN *fringues* ‘clothes’ and the control variant ($p = .012$). Given the high acceptability of this POMN with distributive quantifiers, it seems to be illogical that *fringues* would not be compatible with these highly distributive predicates. One may assume that, just as in the case of test 2a, the colloquial connotation of *fringues* lead to its lower degree of acceptability, but this does not explain the higher mean ratings of *vêtement* in this context.

With respect to the semantic domain of CATTLE, the results are similar:

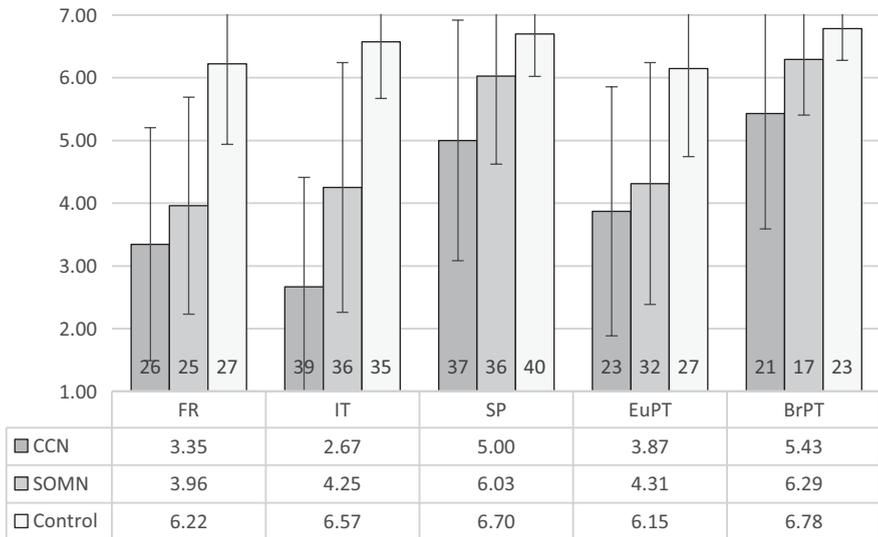


Figure 4.10: Mean acceptability rating of animate Romance OMNs and CCNs in combination with highly distributive predicates.

I again conducted a one-way ANOVA with the factor *condition* for each language and again the differences are significant in every language studied here (*Welch's* $F_{FR}(2, 47.86) = 25.58, p = .000, \eta_p^2 = .512$; *Welch's* $F_{IT}(2, 64.40) = 80.08, p = .000, \eta_p^2 = .713$; *Welch's* $F_{SP}(2, 59.34) = 14.37, p = .000, \eta_p^2 = .327$; *Welch's* $F_{EuPT}(2, 49.15) = 13.85, p = .000, \eta_p^2 = .360$; *Welch's* $F_{BrPT}(2, 29.85) = 6.25, p = .005, \eta_p^2 = .295$). In comparison to the domain of CLOTHING the calculated ANOVAs show somewhat higher effect sizes. A Games-Howell post-hoc test reveals that the CCN condition was rated significantly lower than the control condition ($p_{FR} = .000$; $p_{IT} = .000$; $p_{SP} = .000$; $p_{EuPT} = .000$; $p_{BrPT} = .011$), and these results pattern with the results of the inanimate domain. In Spanish and Italian, the CCN condition was also rated significantly lower than the SOMN condition ($p_{IT} = .002$; $p_{SP} = .032$), while the

SOMN condition, in turn, was rated significantly lower than the control condition in French, Italian, Spanish and European Portuguese ($p_{FR} = .000$; $p_{IT} = .000$; $p_{SP} = .034$; $p_{EuPT} = .000$). The fact that at least in Spanish and Italian the CCN condition was rated significantly lower than the SOMN condition may point to the conclusion that animacy correlating with saliency has an influence on accessibility of the internal plurality of a collection – in the case of the inanimate condition, there were no significant differences between these two collection types. The results for the semantic domain of PEOPLE, however, do not support this. As in test 1b and test 2a, BrPt. *gente* was not included in this test. Figure 4.11 shows the results for this last test domain:

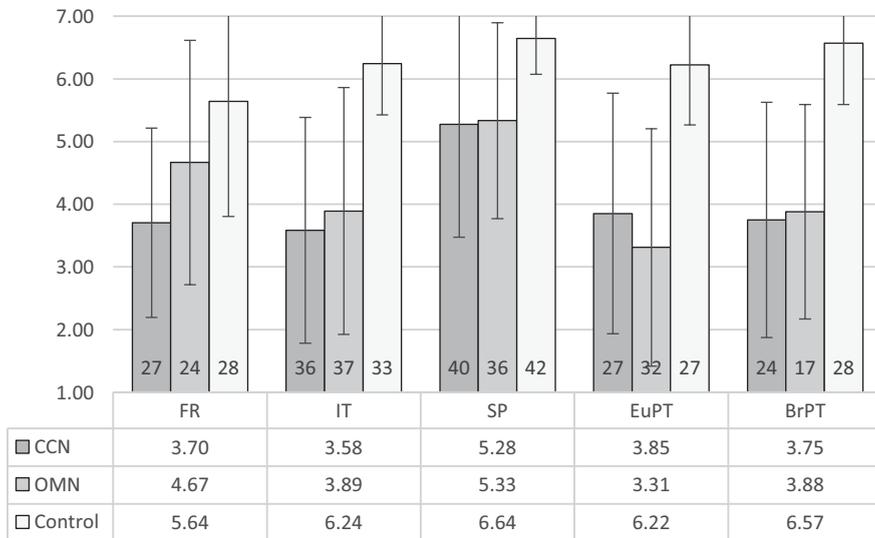


Figure 4.11: Mean acceptability rating of human Romance OMNs and CCNs in combination with highly distributive predicates.

A one-way ANOVA with the factor *condition* again reveals significant differences between the conditions tested for each of the languages examined ($F_{FR}(2, 76) = 7.94, p = .001, \eta_p^2 = .173$; *Welch's* $F_{IT}(2, 61.09) = 44.53, p = .000, \eta_p^2 = .593$; *Welch's* $F_{SP}(2, 58.03) = 19.23, p = .000, \eta_p^2 = .399$; *Welch's* $F_{EuPT}(2, 50.43) = 36.07, p = .000, \eta_p^2 = .589$; *Welch's* $F_{BrPT}(2, 33.16) = 31.46, p = .000, \eta_p^2 = .655$). A Bonferroni post-hoc test revealed that there are no significant differences between the POMN in French and the control condition ($p = .166$), refuting the hypothesis of Cappeau/Schnedecker (2014b, 3037) that the mass character of *gens* blurs the plurality of its constituting individuals. The CCN was rated significantly

lower than the control condition ($p = .000$), but there are no significant differences between it and the POMN *gens* ($p = .183$). A Games-Howell post-hoc test also shows slightly different tendencies in the other languages: the CCN condition is always rated significantly lower than the control condition ($p_{IT} = .000$; $p_{SP} = .000$; $p_{EuPT} = .000$), but this is also the case with the SOMN condition ($p_{IT} = .000$; $p_{SP} = .000$; $p_{EuPT} = .000$), contrary to the French POMN. There are no significant differences between the CCN condition and the SOMN condition ($p_{IT} = .771$; $p_{SP} = .988$; $p_{EuPT} = .541$).

Summarising the results of test 2b for all semantic domains being investigated here, the following can now be stated: in every semantic domain and in every language under study, the CCN condition was rated significantly lower than the plural object noun control condition. This confirms the assumption that the external boundedness of CCNs hinders access to their referential internal plurality (cf. HP_{ss}1). This semantic feature is susceptible to combination with highly distributive predicates like ‘one after the other’, but also with other distributive predicates like stubbornly distributive predicates (cf. test 2a). This feature is not prone to cross-linguistic or other kinds of variation for it could be proven for all semantic domains and languages under study. It was furthermore assumed, following Rothstein (2010a) as well as Pires de Oliveira/Rothstein (2011a), that the combination of an OMN with highly distributive predicates like ‘one after the other’ is susceptible to cross-linguistic variation and that the acceptability of these kinds of construction correlates with the marking of the mass-count distinction of a language. Considering the individual semantic domains, the results were rather heterogeneous. In the human domain, the OMN (singular and plural) was rated significantly lower than the control condition in all the languages examined. In the domains of CATTLE and CLOTHING, however, this did not hold for all languages and the correlations and tendencies determined did not exactly match the assumed continuum of the marking of the mass-count distinction. To address this issue more globally, I summarised the results of all semantic domains, I again excluded BrPt. *gente* for its degree of pronominalisation, Fr. *fringues* for its diaphasic markedness and Fr. *gens* since this would then be the only POMN. To directly compare the SOMNs with the CCNs under study, I additionally included them here. Figure 4.12 shows the summarised mean acceptability ratings of Romance collection nouns in combination with highly distributive predicates:

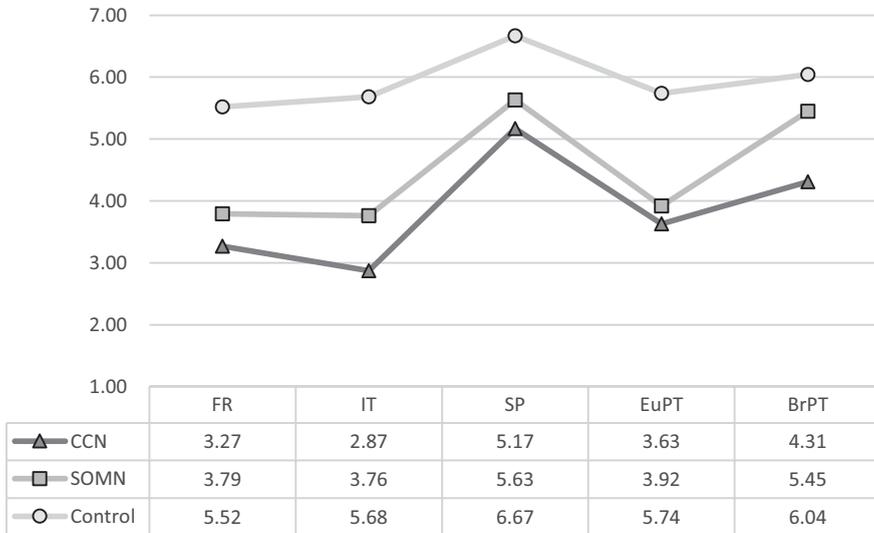


Figure 4.12: Mean acceptability rating of Romance SOMNs and CCNs in combination with highly distributive predicates.

To analyse the global differences between the tested nominal types, I again conducted a one-way ANOVA with the factor *condition* for each of the five languages being investigated. Just as in the case with the separate analyses, there are significant differences between the three conditions in all languages ($F_{FR}(2, 153) = 23.20, p = .000, \eta_p^2 = .232$; *Welch's* $F_{IT}(2, 212.44) = 72.61, p = .000, \eta_p^2 = .294$; *Welch's* $F_{SP}(2, 177.89) = 48.78, p = .000, \eta_p^2 = .161$; *Welch's* $F_{EuPT}(2, 157.79) = 36.21, p = .000, \eta_p^2 = .203$; *Welch's* $F_{BrPT}(2, 81.72) = 8.98, p = .000, \eta_p^2 = .132$). A Bonferroni post-hoc test for French and a Games-Howell post-hoc test for Spanish and European Portuguese reveals that, in these three languages, the control condition was rated significantly better than the SOMN ($p_{FR} = .000$; $p_{SP} = .000$; $p_{EuPT} = .000$) as well as the CCN ($p_{FR} = .000$; $p_{SP} = .000$; $p_{EuPT} = .000$), but that there are no significant differences between the latter two categories ($p_{FR} = .406$; $p_{SP} = .122$; $p_{EuPT} = .603$). Only in the case of Italian and Brazilian Portuguese was the CCN condition rated significantly lower than the SOMN condition ($p_{IT} = .001$; $p_{BrPT} = .030$). In addition, the SOMN condition was also rated significantly lower than the control condition in Italian ($p = .000$), but there are no differences between the SOMN and the control condition in Brazilian Portuguese ($p = .249$). This confirms the generalisations that CCNs are not compatible with highly distributive predicates, and that this incompatibility is not susceptible to cross-linguistic variation. With regard to the acceptability of the tested SOMNs, there is the generalisation that these are less acceptable in French, Italian, Spanish and

European Portuguese, but as acceptable as the control condition in Brazilian Portuguese. In addition, there is no difference between the CCN and the SOMN in French, Spanish and European Portuguese, but the SOMN was rated significantly higher in Italian and Brazilian Portuguese. Considering these results, one can summarise that the assumed continuum of acceptability of SOMNs with highly distributive predicates should be narrowed to an opposition between Brazilian Portuguese on the one hand and the other tested languages on the other. For this test domain, the second semantic-syntactic hypothesis (HP_{ss2}) can only be confirmed to a certain extent: there seems to be a systematic correlation between the acceptability of highly distributive predicates in combination with SOMNs and the degree of grammaticalisation of the mass-count distinction in the Romance languages, but only in a dichotomous way – Brazilian Portuguese which does not overtly mark the mass-count distinction vs. the other Romance languages which mark it to at least some degree.

To shed a little more light on possible differences between morphologically singular versus plural OMNs, an additional set of data can be included here. In the context of the main study, I also tested some other OMNs in combination with this highly distributive predicate. These test sentences were rather exploratory and not as systematic as the ones presented above. In Italian, a direct comparison between the POMN *stoviglie* ‘dishes’ and the SOMN *vasellame* ‘tableware’ was made. This Italian test sentence tested the combination of <‘dishes/crockery’ + ‘one on top of each other’>:

- (49) Alessandro non ha assolutamente posto nella sua credenza. Deve mettere <le stoviglie_{POMN}/il vasellame_{SOMN} sempre *una/o sull'altra/o*>.
 ‘Alessandro has absolutely no space in his kitchen cupboard. He always has to put <the dishes/the crockery on top of each other>.’

An independent two-sample t-test reveals that the SOMN is rated significantly lower than the POMN ($M_{SOMN} = 4.28$; $M_{POMN} = 5.61$): $t(75) = 3.43$, $p = .001$, $d = .08$. Although this is only a small addition to the puzzle, it indicates that the incongruent results in French may indeed be due to the connotation of *fringues* and that the internal plurality of POMNs is – as has been hypothesised – more easily accessible than that of a SOMN. It could, however, also be the case that *stoviglie* is just more common than *vasellame*, as the comments in the context of test 2a suggest.

A last issue which has to be addressed concerns possible animacy effects of the acceptability of the construction <‘one after the other’ + OMN>. As assumed according to the third semantic-syntactic hypothesis (HP_{ss3}), there should be an

increasing accessibility of the internal plurality of an OMN with an increasing degree of animacy of the constituting referents. To analyse this issue, I conducted a two-way ANOVA with the factors *condition* and *degree of animacy*. With respect to the condition, I only included the SOMNs (again except for BrPt. *gente*) since both the state of the art and the results of these empirical examinations confirm an incompatibility of CCNs with highly distributive predicates. The POMNs are also excluded for the issues of incongruency described. The analysis of variance reveals a large main effect on the factor *condition* ($F(1, 804) = 165.81, p = .000, \eta_p^2 = .171$), but only a small main effect on the factor *degree of animacy* ($F(2, 804) = 21.28, p = .000, \eta_p^2 = .050$) and only a slight interaction between the two factors ($F(2, 804) = 8.44, p = .000, \eta_p^2 = .021$). A Bonferroni post-hoc test reveals significant differences between all three degrees of animacy (inanimate vs. animate: $p = .000$; animate vs. human: $p = .021$; inanimate vs. human: $p = .003$), but these show nearly the opposite of what has been assumed: in the human domain, the difference between SOMN and control condition ($M_{SOMN} = 4.21; M_{Cont.} = 6.40$) is higher than in the case of the animate condition ($M_{SOMN} = 4.89; M_{Cont.} = 6.50$), but lower in comparison with the inanimate condition ($M_{SOMN} = 4.32; M_{POMN} = 5.25$). The animate SOMNs were thus rated comparably lower than the inanimate SOMNs. This means that the SOMNs tested were most acceptable in the inanimate domain of CLOTHING, they show an intermediate acceptability in the animate domain of CATTLE, and they are least acceptable in the human domain of PEOPLE. Consequently, a systematic effect on the degree of animacy cannot be confirmed.

Summarising the overall results of tests 2a and 2b, which examined the accessibility of the referential internal plurality of a collection noun in accordance with the hypotheses formulated at the beginning of chap. 4, the following can now be stated: the cross-linguistic variation assumed on the basis of the differences indicated between English and Brazilian Portuguese by Rothstein (2010a) and Pires de Oliveira/Rothstein (2011a) could not be confirmed. The tested Romance OMNs were all less acceptable than the plural object noun control variant and there was no considerable interaction between the factors of condition and language. There is, however, still the issue of the good acceptability of BrPt. *roupa* to be solved. It was assumed in the framework of tests 1a and 1b that it is perhaps only members of the domain of artefactual OMNs that are more acceptable in these kinds of syntactic environment. One last piece of evidence, however, indicates that it seems to be the single lexeme of *roupa* which is aspectually flexible, not the whole class of Brazilian Portuguese SOMNs. In the framework of the more exploratory testing of some other OMNs in combination with highly distributive predicates, the semantic domain of FURNITURE was also tested. The tested sentence was the following (the Portuguese version is given for its relevance in the present context):

- (50) Depois da festa puseram <o mobiliário_{SOMN}/a mobília_{SOMN} um/a sobre o/a outro/a> para limpar a sala superficialmente.
 ‘After the party, they put <the furniture one on top of each other> to clean the room superficially.’

Since these sentences were only exploratory, there are no systematic control and test sentences, but the merely mediocre evaluations of all nouns in every language under study suggests that the good acceptability of Pt. *mobília* indicated in Rothstein (2010a) and Pires de Oliveira/Rothstein (2011a) is probably indeed due to the prepositional phrase *dessa marca* ‘of this brand’, and not necessarily to a generally better accessibility of the internal plurality of a mass collection in Brazilian Portuguese (cf. Figure 4.13).

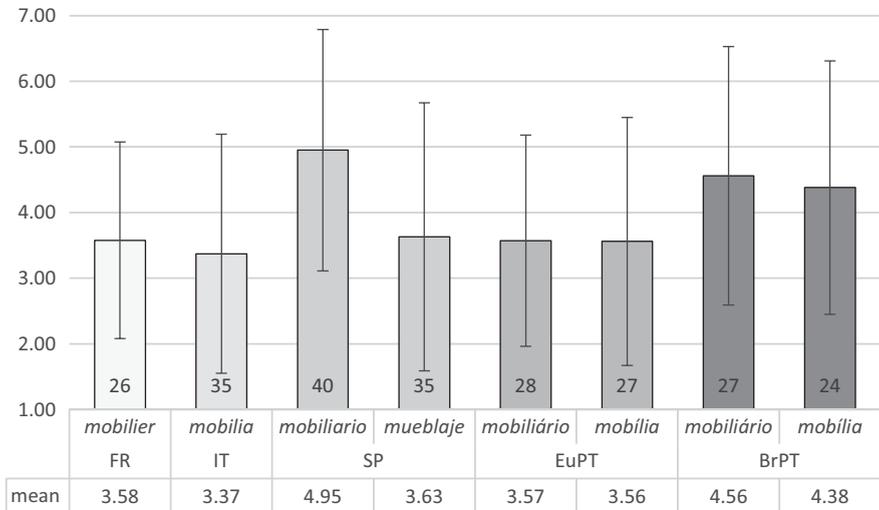


Figure 4.13: Mean acceptability rating of Romance ‘furniture’-denoting SOMNs with highly distributive predicates.

The cross-linguistic variance as assumed in the second semantic-syntactic hypothesis (HP_{ss2}), thus could only be confirmed for the semantic domain of CLOTHING and only for the syntactic feature of countability. With respect to the first semantic-syntactic hypothesis (HP_{ss1}), tests 2a and 2b evidenced considerable differences between the tested nominal types. Both tests confirmed an absolute unacceptability of bounded CCNs with stubbornly distributive predicates and other highly distributive predicates. Test 2a additionally confirmed the assumed good compatibility of morphologically singular and plural OMNs with stubbornly dis-

tributive predicates, but test 2b also showed a significantly lower acceptability of those nominal types with highly distributive predicates like ‘one after the other’. An explanation for these differences could be that a stubbornly distributive predicate, although it refers to the single individual entities of an OMN, still leaves open its actual quantity – adjectives like *round*, *small* or *long* do not predicate a specific, contextually determined and exact number of referents, but the indefinite mass that an OMN predefines. Specifically, *long clothing* or *rectangular dishes* do not mean that every single piece of clothing and tableware is long or rectangular, but only that some items may be predicated as such, neglecting e.g. the underwear or the wine glasses, so this actual predication is left unexpressed. In the case of constructions like ‘one after the other’ this kind of under-specification cannot be maintained: putting furniture on top of each other or clothing one after the other in the cupboard predefines a very specific set of entities not available in the denotation of an OMN. For the same reason that we cannot count OMNs, we cannot – or can hardly – combine them with this kind of highly distributive predicate, irrespective of the nominal system of a language. The question regarding whether there are differences in this respect between morphologically singular and plural OMNs could not be addressed systematically. The good acceptability of Fr. *gens l’un après l’autre* ‘people one after the other’, however, indicates that the morphological form indeed facilitates the combination with highly distributive predicates. POMNs thus semantically link their constituting referents by maintaining some kind of distributivity, also displayed by their combinability with distributive quantifiers. With respect to the third semantic-syntactic hypothesis (HP_{ss}3), test 2b had to refute the hypothesis of an increasing accessibility of the internal plurality of an OMN with an increasing degree of animacy. This is supported by the fact that test 1b did not show any differences between Fr. *gens* and *fringues* with respect to their acceptability with more or less distributive quantifiers. Just as in the case of the mostly impossible constructions ad sensum in Romance languages (cf. chap. 3.1.2), there are no animacy effects in the case of distributive quantifiers and determiners.

Test 2c: The last test domain which will be addressed in this empirical section concerns the hierarchical meaning relations between the collection itself and its constituting members. In chap. 3.1.2, it was assumed in this respect that SOMNs are characterised by both hyponymic as well as meronymic relations, but that they do not represent good exemplars for hyperonyms nor holonyms. It was additionally assumed that POMNs may be rather part of a meronymic than of a hyponymic hierarchy in terms of being determined by contiguous functionality. In the following remarks, I will first present some further considerations on OMNs in

hyponymic and meronymic relations and will then present the results for both tests together.

OMNs are not good candidates for a typical taxonomy, since the similarity-relation between the superordinate term and its subordinate entities does not lie in a specification or generalisation of a core property, a prerequisite for taxonomies (cf. Cruse 2011, 137). In contrast, the actual kind of similarity seems to vary from noun to noun, where sometimes it is the common function (cf. e.g. *furniture*), and sometimes certain common core properties (e.g. perceptual) that can ultimately be relevant for classification (cf. e.g. *fruit*; cf. also Grimm/Levin 2016). For these reasons, I did not test a possible hyperonym status of the different SOMNs by means of the canonical construction *X is a type of Y*, but by means of *X and other Ys*. This latter construction may reveal class inclusion relations in a far less restricted manner and it is also possible to include the feature of heterogeneity in this construction – a crucial aspect for OMNs: “[. . .] it is possible that for a pair of words A and B, some, but not all, construals of A may be hyponyms of some, but not all construals of B” (Cruse 2011, 135). Since I do not expect any cross-linguistic differences in this semantic test domain, I differentiate in what follows between SOMNs and POMNs, focusing (similarly to the test on stubbornly distributive predicates) on French and also on Spanish. I do not differentiate between the different degrees of animacy either, since I am not expecting any effects on the rating in this test domain. The control condition in this set of test sentences is a plural object noun (PON) on the superordinate level, like Fr. *personnes* ‘persons’. Since I want to include the factor of heterogeneity in this test domain, I did not test only one single X, but a small heterogeneous list of Xs like Fr. *mon frère, un collègue et. . .* ‘my brother, a colleague and . . .’. Example (51) shows the sentences tested for French and Spanish:

- (51) a. FR: Hier, j’ai acheté <une chemise, un manteau *et d’autres fringues*_{POMN}/
*d’autre vêtement*_{SOMN}/*d’autres vêtements*_{PON}>.
 SP: Ayer compré <una camisa, un abrigo y *otra ropa*_{SOMN}/*otras prendas*_{PON}>.
 ‘Yesterday, I bought <a shirt, a coat and other clothes/clothing/pieces of clothing>.’
- b. FR: À la ferme de ma tante, il y a <deux chevaux, une vache *et d’autre bétail*_{SOMN}/*d’autres bêtes*_{PON}>.
 SP: En la granja de mi tía hay <dos caballos, una vaca y *otro ganado*_{SOMN}/*otros animales*_{PON}>.
 ‘At the my aunt’s farm, there are <two horses, a cow and other cattle/animals>.’

- c. FR: Hier, j'ai rencontré <mon frère, un collègue *et d'autres gens*_{POMN}/*d'autres connaissances*_{PON}>.
 SP: Ayer me encontré con <mi hermano, un colega y *otra gente*_{SOMN}/*otras personas*_{PON}>.
 'Yesterday, I met <my brother, a colleague and other people/acquaintances, persons>.'

As indicated in chap. 3.1.2, collections represent – if at all – a special type of holonym. The focus in these kinds of meronymic relation does not lie on the whole, as in the case of a bicycle that has various parts, but on the entities making up the collection. Following Lecolle (1998, 52–53), a definitional construction for the former type is e.g. Fr. *X a Ys* 'X has Ys' or *X comporte Y* 'X comprises Y'; these are diagnostic for meronymic relations of the type *whole* > *parts*. For OMNs, which are of the opposite type, namely *parts* > *whole*, definitional constructions are Fr. *X se compose de Y* 'X makes up Y' or *X est composé de Y* 'X is made up of Y'. For these reasons, I test the meronymic properties of OMNs in the Romance languages focused on here with exactly these latter constructions. The OMNs tested are determined in these contexts by a possessive determiner. Since these kinds of collection are not wholes that have stable, expectable parts (like a bicycle), a context that creates these expectations is necessary. I thus constructed for each sentence a contextualising introduction followed by the testing construction. The control variant in this set of test sentences is a CCN. As in the case of the set that tested the hyponymic relations, I made sure that the constituting entities represent a heterogeneous set. Example (52) shows the test sentences for French and Spanish:

- (52) a. FR: Aujourd'hui, Amandine est joliment vêtue, <ses fringues_{POMN}/son vêtement_{SOMN}/sa tenue_{CCN} *se compose(nt) d'un chemisier, d'une jupe et d'un foulard*>.
 SP: Hoy Inés está bien vestida, <su ropa_{SOMN}/su atuendo_{CCN} *se compone de una blusa, de una falda y de un pañuelo*>.
 'Today, Amandine is well dressed, <her clothes/her clothing/her outfit consists of a blouse, a skirt and a scarf>.'
- b. FR: L'agriculteur Deschamps a beaucoup d'animaux différents, <son bétail_{SOMN}/son troupeau_{CCN} *se compose de chevaux, de bœufs et de moutons*>.
 SP: El agricultor Sánchez tiene muchos animales diferentes, <su ganado_{SOMN}/su rebaño_{CCN} *se compone de caballos, cabras y ovejas*>.
 'The farmer Deschamps has many different animals, <his cattle/his herd consists of horses, cows and sheep>.'

- c. FR: Demain la course annuelle d'entreprises aura lieu, <nos gens_{POMN}/notre équipe_{CCN} se compose(nt) seulement d'hommes et de femmes jeunes>.
 SP: Mañana tendrá lugar la carrera anual de las empresas locales, <nuestra gente_{SOMN}/equipo_{CCN} solamente se compone de hombres y mujeres jóvenes>.
 'Tomorrow, there will be the anual company run, <our people/our team consists of only young men and women>.'

Against the backdrop of the state of the art described in chap. 3.1.2, my hypothesis for this last test domain is that there should be no significant differences between the SOMNs and POMNs tested, but significant differences between the control condition and the tested OMNs are possible. Figure 4.14 shows the results for both tests examining the hierarchical relations between the collection and its constituting entities, with French and Spanish taken together (the number of evaluations considered is again depicted at the bottom of each column).

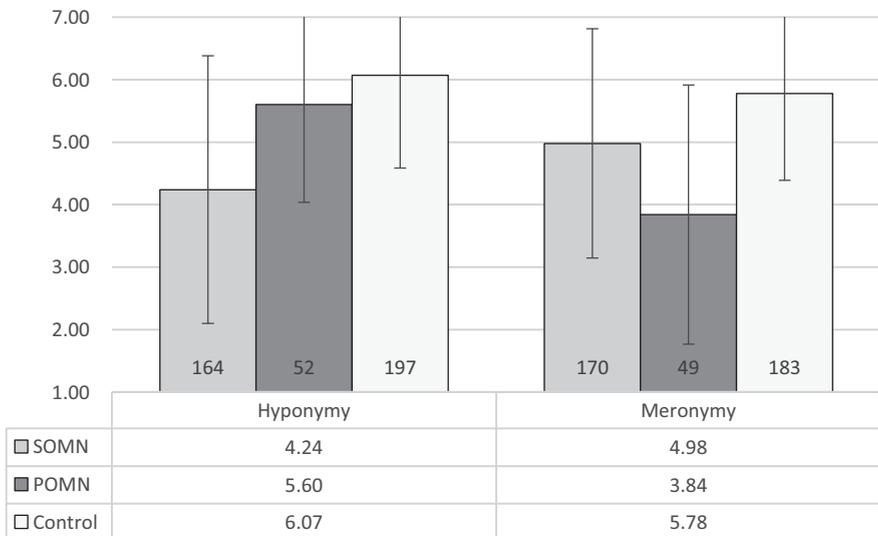


Figure 4.14: Mean acceptability rating of Romance OMNs in hyponymic and meronymic constructions.

In the hyponymic testing domain, a one-way ANOVA with the factor *condition* reveals significant differences between the three aspectual types, and the category of POMNs is again represented only by French *gens* and *fringues* (*Welch's* $F(2, 143.39) = 42.28$, $p = .000$, $\eta_p^2 = .189$). A Games-Howell post-hoc test additionally shows that the SOMN condition was rated significantly lower than the control con-

dition ($p = .000$) and the POMN condition ($p = .000$). It can thus be confidently confirmed that SOMNs are not good hyperonyms, but they are not impossible in these constructions either: the mean rating of 4.24 is intermediate rather than low. For the two POMNs *gens* and *fringues* in French, there is no significant difference between them and the control condition ($p = .132$) – the hypothesis of Wierzbicka (1984, 321–322) that this aspectual type is characterised rather by a meronymic contiguity relation and not by a hyponymic similarity relation thus cannot be confirmed. It may be, however, that the reason for this is simply because Wierzbicka mostly considered POMNs like *groceries* which are still semantically very close to ad hoc categories like *things to take on a camping trip*. This does not hold for the intension of *gens* and *fringues*. The plural morphology may have improved the hyponymic construction.

Another one-way ANOVA also reveals significant differences between the three aspectual types in the testing domain of meronymic relations (*Welch's* $F(2, 124.15) = 25.01$, $p = .000$, $\eta_p^2 = .126$). A Games-Howell post-hoc test shows for this testing that the control condition was rated significantly better than the SOMNs ($p = .000$) and the POMNs ($p = .000$). In addition, the POMN condition was now rated as significantly less acceptable than the SOMN condition ($p = .002$). It can once more be confirmed that SOMNs are not good holonyms, but again it is not impossible for them to function as such. The empirical results thus confirm the (theoretical) assumption of Winston/Chaffin/Herrmann (1987, 428) that “a chair is both a kind of furniture [. . .] and an item of furniture”. The explanation for the lower acceptability of the two POMNs in this meronymic construction has already been given above: the two specific nouns tested in French are simply not constituted by contiguity, and it may well be that other POMNs would have shown better results under this test condition.

The aim of this chapter was to empirically examine the semantic-syntactic properties of French collection nouns in comparison to Spanish, Italian and Portuguese. The superordinate research question (RQ1) was: what are the influencing factors on the particular linguistic expression of a collection of entities and the semantic-syntactic characteristics related to it? After this description and discussion of the results of the empirical studies conducted for this study, the following chapter will summarise the results against the backdrop of the formulated hypotheses in order ultimately to refine the preliminary model of nominal aspectuality presented in chap. 3.1.3

4.2.3 Revising the integral semantic-syntactic model of collections

The integral semantic-syntactic model of collection noun types as suggested in Figure 3.1 started from the premise that the aspectual constitution of a noun, i.e.

the construal of a plurality of entities verbalised by a certain noun, is defined by the two parameters of external boundedness and internal plurality. It was assumed in this respect that these primarily conceptual features correlate with the linguistic properties of a noun. An increase in the set profiling of a collection thus goes in hand with an increase in the ease of countability of the noun denoting it. It was also assumed that the ontological type of referents influences their accessibility by linguistic means, i.e. that collection nouns referring to human referents are rather combined with e.g. *construiones ad sensum*. As also discussed, the model was, however, mainly based on the state of the art on French CCNs and English OMNs, which in turn was mostly based on introspection. The empirical study presented in this chapter consequently aimed primarily to put these mainly theoretical assumptions on an empirical foundation.

The first research question (RQ_{ss}1) concerned possible differences between the three main nominal types of CCNs, SOMNs and POMNs. Given some first indications that e.g. CCNs are not compatible with adjectives predicating the individuals of a collection, the question in this respect addressed the extent to which morpho-syntactic features like countability and morphological number correlate with semantic aspects like the constitution of the internal plurality of a collection or the possible linguistic accessibility of it. In concrete terms, I assumed CCNs not to be compatible with any kind of linguistic means that accesses their internal plurality, but SOMNs to be acceptable with these means and POMNs to perhaps even be an ideal fit as plural object nouns with distributive predicates. I therefore suggested that the fossilised morphological number does indeed play a role for the acceptability of an OMN with a distributive predicate. The acceptability judgement tests overall confirmed the absolute incompatibility of CCNs with any kind of distributive predicate, and this was true for all languages and all semantic domains investigated. The acceptability judgement tests also confirmed the assumed compatibility with OMNs, irrespective of the morphological number, with stubbornly distributive predicates. With respect to combinability with highly distributive predicates like ‘one after the other’, the acceptability judgement tests additionally confirmed differences between the tested SOMNs and POMNs, which means that the morphological form either facilitates or complicates the combination with a highly distributive predicate. Test 2c moreover showed that OMNs differ in the hierarchical constitution of their internal pluralities. In this respect, SOMNs were both acceptable in meronymic and hyponymic contexts, but not perfect. This confirms that SOMNs may be used as hyperonyms and holonyms, but also that they are not good candidates for these functions. POMNs like Fr. *gens* were more acceptable in hyponymic than in meronymic contexts, which brings them closer to count superordinates.

The second research question (RQ_{ss2}) examined the extent to which linguistic characteristics of English OMNs also hold for Romance languages and the degree to which the language-specific nominal system influences these characteristics (cf. Table 1.2). Based on the state of the art summarised in chap. 3.1, the hypothesis concerning this question was twofold: on the one hand, I assumed Romance OMNs to behave basically the same as English OMNs, since English and the Romance languages are all of the same nominal number type in differentiating between ‘one’ and ‘more than one’. On the other hand, the differences between the various Romance languages in their marking of mass and count nouns led to the assumption that Romance OMNs may indeed differ in their linguistic properties and therefore also differ from English. These differences were mainly assumed to be present when it comes to the syntactic marking of mass nouns, i.e. their incompatibility with the indefinite article as well as their incompatibility with the inflectional plural in the case of SOMNs. The examination of the compatibility of Romance OMNs meaning ‘clothing’ with the indefinite article led to a resulting continuum of accessibility of such a construction going from French to Spanish, European Portuguese, Italian and finally Brazilian Portuguese. This continuum consequently does not correspond exactly to the assumed continuum based on considering the different nominal systems marking the mass-count distinction to a lesser or higher degree. It. *abbigliamento* did not fit the picture here, but since it was significantly more acceptable with verbs implying a bounded collection, it was assumed that *abbigliamento* still retains much of its original meaning of ‘attire’. The hypothesis was thus confirmed that combinability with the indefinite article indeed correlates with the degree of grammaticalisation of the mass-count distinction. This correlation could not, however, be confirmed in the case of the inflectional plural. A comparison between a pluralised SOMN and plural object noun as a control variant led to the conclusion that the acceptability of the inflectional plural does not correlate with the nominal system of the language, viz. the obligatoriness and overtness with which it distinguishes mass from count nouns. A possible explanation for these incongruencies might be that the two syntactic means of expression of countability are simply not equally susceptible to these correlations. In this sense, there might be differences between singular and plural countability. The good acceptability of Pt. *roupa* ‘clothing’ with both the indefinite article and the inflectional plural, however, leads rather to the assumption that the described correlations may only be valid in cases of artefact nouns and maybe only in the case of nouns denoting ‘clothing’. Future research should thus focus on a systematic comparison of various ontological types of OMN with both the indefinite article and the inflectional plural. Test 2b of the acceptability judgement study furthermore refuted the hypothesis of cross-linguistic variation in constructions accessing the internal plurality of a collection noun. Whereas

there were differences between the tested nominal types (cf. *supra*), at least in the investigated Romance languages and in the case of the tested nouns, there was no correlation between the acceptability of constructions like ‘one after the other’ and the nominal system of the language – they were all equally (un-)acceptable. The consideration of exploratory sentences testing the various Romance equivalents of Engl. *furniture* additionally points to the assumption that it seems to be only BrPt. *roupa* that has a special status amongst the tested collection nouns.

The final research question addressed by the acceptability judgement tests (RQ_{ss3}), was the possible influence of the ontological type of the referents making up the collection. It was assumed, given the various phenomena related to the animacy hierarchy, that the degree of animacy of the referents correlates with the degree of accessibility of them. Neither of the tests could confirm this hypothesis. There were no differences between the inanimate, animate and human collection nouns in the tests on the plural inflection, and the detected differences between the three semantic domains in the tests on highly distributive predicates did not match the animacy hierarchy. The tested phenomena thus behave like the constructions *ad sensum* which were – as has been shown – also nearly impossible in the Romance languages studied, in contrast to English.

The empirical tests discussed in chap. 4 lead to a revised integral semantic-syntactic model of collection nouns in Romance languages. The main assumptions could be confirmed, this applies particularly to the differences between the three main aspectual types of CCNs, SOMNs and POMNs. The study, however, refuted the hypothesis of differences depending on the animacy of the referents, and only some of the cross-linguistic correlations could be confirmed. Following on from this, it may be assumed that the acceptability of (lexical) means of accessing the internal plurality of a collection noun, *viz.* stubbornly distributive and highly distributive predicates, does not depend on the nominal system of a language, whereas the acceptability of those means that access the outer boundaries, thus the indefinite article, may be language-dependent. This goes in line with the generalisations made in chap. 1.1, that linguistic means of expression of boundedness are cross-linguistically prone to variation, but linguistic means of expression of internal structure are more or less language-independent – tables and chairs may be syntactically coded differently, but predicating them as round and square should not depend on the specific linguistic system. Furthermore, the study in general supported an important consideration: research on collection nouns, especially in the case of SOMNs, has normally been conducted from the perspective of the English language, often implying that the same features also apply in other languages. The present study allowed to uncover significant differences between various languages, but also between various specific nouns. This points to the fact that general tendencies for collection nouns may well be determined,

but idiosyncratic features of specific nouns or noun types should not be underestimated when analysing their linguistic and presumably underlying conceptual features. This supports the usage-based approach adopted in the framework of this present study. This is illustrated by the results of the acceptability of SOMNs with the inflectional plural. There were no systematic cross-linguistic differences in the acceptability of the tested SOMNs with the inflectional plural, with only Pt. *roupa* ‘clothing’ being significantly more acceptable than the other tested variants, while the effect was, however, not present in other tested nouns. This suggests that Pt. *roupa* alone is flexible in its use, since it can be used as a SOMN, as a CCN or as an object noun in the singular and plural. With this, it may to some extent replace the uncomfortable countable construction *peça de roupa* ‘piece of clothing’. For instance, there are 157,294 occurrences of *roupas* in the *ptTenTen11* corpus, but only 2,309 of *peças de roupa*. Spanish and Italian both have an equivalent count superordinate that is not a long binominal construction but a simple word (*prenda* and *vestito*, respectively). Figure 4.15 represents, summarising the main aspects of chap. 4, the revised schematic model of collection nouns in Romance languages:

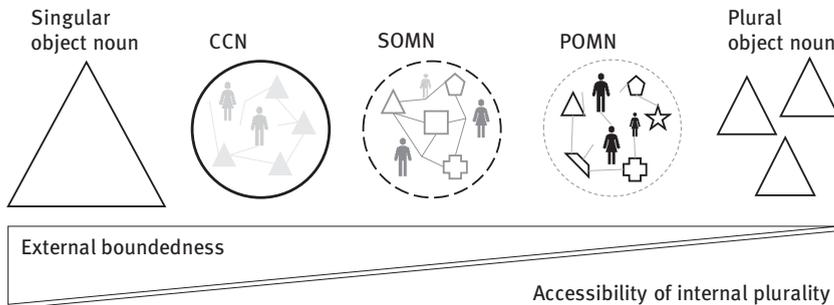


Figure 4.15: Revised integral semantic-syntactic model of Romance collection nouns.

Against the background of the huge amount of research on the topic of the mass-count distinction, the interpretation of OMNs in these respective theoretical frameworks as well as the language comparisons made by a number of authors (mostly English vs. Mandarin Chinese; cf. e.g. Chierchia 1998a), one might ask what the theoretical implications of this revised model of Romance collection nouns would be. As discussed in this chapter, some of the various assumptions could be confirmed with the results of the acceptability judgement study, some generalisations were specified or restricted. In addition to the different approaches to the mass-count distinction discussed in chap. 3.1.1, there is the so-called exo-skeletal approach of Borer (2005, 93–94) who assumes that “all nouns, in all languages, are mass [. . .] and that mass interpretation is, in a sense to be defined, a default

interpretation, associated with the absence of a dividing structure” (cf. also Bale/Barner 2009 for a similar approach). According to this view, it is thus the syntactic structure a root noun is inserted in which determines its syntactic properties and not the semantics or lexically inherent properties. The results of the study presented in this chapter refute this approach in two ways: On the one hand, the results show that countability is no dichotomous operation, but rather has to be seen as a continuum. Consequently, some nouns and also some determiners are more or less countable than others (cf. Allan 1980 for a similar view). On the other hand, the various degrees of acceptability of certain constructions, but also the relatively high degree of variation in the evaluations indicate that nouns simply are not flexible. That is to say: the countability preferences of a noun are inherent to it and they may or may not predetermine the possibility for kind readings (*three wines* vs. **three sands*) or ambiguities between e.g. a SON- or an OMN-reading. These properties are specific not only to different languages, but even to different language varieties, as shown by the comparison between European and Brazilian Portuguese. Language may thus be considered a tool which is not only shaped by general cognitive processes like entity construal, but also by language use itself, world knowledge or – as will be shown in part III. – by diachronic evolution.

5 Morphological characteristics: A corpus analysis of collective nonce-formations

The main result of chap. 3 and the overall recapitulation of the state of the art was that the category of collection nouns in Romance languages with its diverse subtypes can best be defined through the semantic-syntactic aspects of countability and the accessibility of the constituting referents. Attempts to also include derivational processes producing collection nouns have mostly been restricted to lexicographic analyses. These studies concluded that collection nouns as a sharply delimited derivational category do not exist, given the high polyfunctionality of collectivising Romance suffixes. In addition to the non-exclusiveness of collective suffixes, many authors have also commented on the low level of productivity of this derivational process. In this respect, Grossmann (2004, 244) e.g. states that “è ridotto anche il numero dei suffissi che sono sincronicamente produttivi” (‘the number of suffixes which are synchronically productive is small’). In most cases, this understanding of productivity is however directed to past evolution of language, although *sincronicamente* refers in this context to present-day language use. Grossmann (2004) as well as e.g. Santiago Lacuesta/Bustos Gisbert (1999) base their evaluation of the productivity of a certain (collective) suffix on consulting dictionaries. Even very recent attempts to approach the topic of derivational processes involved in the forming of collection nouns still stick to this methodology. For instance, Cohen (2020) analyses derivational processes relevant for the forming of OMNs in English, French and Hebrew. As relevant and important as her research question and analysis may be, this study, like the others, is still unable to give more insights into this topic than my own considerations made in chap. 3.2: she states that derivatives on *-age* or *-erie* may represent OMNs, but she summarises by citing Aliquot-Suengas (2003) that “the suffix *-aille* is the only one of the set that produces aggregate-mass terms, while the others produce structured groupings” (Cohen 2020, 50). Still she speaks of “the productivity of aggregate derivation” (Cohen 2020, 52) without defining what exactly her understanding of productivity is – supposedly also past productivity since she only mentioned lexicalised derivatives. This methodology of only considering already existing derivatives mostly on the basis of lexicographic information implies, however, several problematic aspects, such as the representativeness of the examples. Given e.g. the commercial and didactic purpose of a dictionary, it often displays only the common, frequent forms a user may want to look up. In addition, dictionaries are only able to give an overview of types, not of tokens (cf. Gaeta/Ricca 2003, 63–65). Following Baayen (2009), this kind of productivity is called *realised productivity*, the number of types a derivational process has already formed. In addition to the problematic issues associated with basing the evaluation of a deriva-

tional process only on realised productivity, it equally cannot answer the question of whether certain collection noun sub-types are derived via special derivational processes. That is to say, by only considering French collective derivates on *-erie*, one might state that they are numerous and that many also represent OMNs (e.g. Fr. *lingerie* ‘underwear’), but it is not possible to say whether this process directly derives mass collection nouns or whether these only came into being by diachronic evolution. This question will amongst others be addressed in the framework of my own diachronic analysis in chap. 8. In contrast to the existing research on collection nouns as a derivational category, another approach to productivity, namely its potential to expand a morphological category, might shed some light on the basic question of productivity of Romance collective suffixes and also the question of the derivation of collection noun sub-types, e.g. OMNs. This kind of *expanding productivity* is not based on already lexicalised words but on *hapax legomena*, i.e. words that occur only once in a given corpus. The main assumption here is that productivity is displayed by “the probability of coming across, new, unobserved types” (Baayen/Lieber 1991, 809; cf. also Baayen 1993). In this respect, productivity P is the ratio of hapax legomena n_i derived upon a certain morphological process in relation to the entire number of token N of this morphological category in a given corpus. Despite criticism of this kind of measurement (cf. e.g. Marle 1992), the basic idea of considering hapax legomena to capture the degree of productivity of a derivational process has its advantages over e.g. a mere type-token-ratio calculation: “While a high type-token ratio may indicate a process or reading that gives rise to a high number of types with a low token frequency, which would be expected of a productive process, it may also be the result of a process or reading that contains only a very small number of types with a low token frequency. Such a process or reading could not be seen as overly productive” (Schulte 2015, 62). In what follows, I will therefore focus on the analysis of expanding productivity of collection nouns, concentrating not on only lexicalised derivates, but also on nonce-formations derived with the collective suffixes of Table 3.1. This methodological approach of combining realised with expanding productivity allows a detailed and well nuanced examination of a derivational category, its productivity in present-day language as well as the possibilities and restrictions it is subjected to.

Before coming to a more concrete elaboration of my methodology, a short note on terminology is necessary. As mentioned above, Baayen bases his quantitative approach on expanding productivity on hapax legomena, thus on words that only occur once in the corpus. Strictly speaking, these may or may not be transparent derivations in being interpretable from a researcher’s point of view. Since this present analysis is an onomasiological examination and the derivates found in a corpus should be transparent to enable categorisation, I will not choose in what follows the notion of *hapax* to describe my analysis and the results – at least

not primarily. Neither will I use the notion *neologism*, because this term implies to a certain extent that the word is already lexicalised in that it is found in dictionaries etc. (cf. e.g. *MLS*, s.v. *Neologismus*; cf. also chap. 7.1 for a discussion of the concept of *lexicalisation*). I will instead use the term *nonce-formation*, which will be adopted in the sense of Bauer (1983, 45) who defines it as “a new complex word coined by a speaker/writer on the spur of the moment to cover some immediate need”. I am well aware of the definitional discussion that nonce-formations in the strict sense need not exactly be new to the language community, but only to the individual creating speaker him- or herself (cf. Hohenhaus 2005, 364). Unfortunately, there does not seem to be any English term which exactly describes a unique complex word, derived as a communicative need and which consequently reflects the underlying derivational schema and which can thus effectively indicate the expanding productivity of it (equivalent to e.g. Germ. *Neubildung*). I will thus adopt various terms in their quite broad senses, like *hapax legomenon* (focusing on the uniqueness of the derivate), *nonce-formation* (focusing rather on the creativity involved) or also *ad hoc construct* or *ad hoc formation* (i.e. a derivate created in a specific context for a specific communicative need).

An indication that a consideration of neologisms in general and nonce-formations in particular is fruitful in the domain of collection nouns is given by research on Italian *ad hoc* categories. Mauri (2017) and Magni (2018) investigate linguistic means of expression that denote a spontaneous clustering of a plurality of entities not only in Romance languages (especially in Italian), but also in e.g. Kuuk Thaayorre, an Australian language (cf. Mauri 2017, 309–311). The authors identify various Italian derivational processes that lead to nouns denoting spontaneous collections built around one central orientation point and through either contiguity (X & Co.) or similarity (X & associates). In this respect, Magni (2018) states in relation to the already mentioned Italian suffixes *-ame*, *-ume* and *-aglia* (cf. chap. 3.2) that they are synchronically productive for deriving *ad hoc* categories mostly based on the proper names of politicians.⁵⁰ It is exactly these kinds of process and examples of neologism that are relevant measuring the productivity of collective suffixes in Romance languages. In example (53a), the nonce-formation *berlusconame* denotes a collection based on a contiguity relation between the former Italian prime minister Silvio Berlusconi and people having to do with him, whereas in (53b) the relation of entities making up the collection is based rather on a similarity relation between the former prime minister Matteo Renzi and people like him.

⁵⁰ This special derivational process is not necessarily limited to deriving collections. Huguin (2018) shows for French, that the formation of nouns, verbs as well as adjectives on the basis of proper nouns of politicians is a very productive word formation pattern here.

- (53) a. Dire che la Boldrini è uguale a Mastella, al figlio di Bossi o al *berlusconame* è una violenza ideologica che non porta da nessuna parte. (Mauri 2017, 310)
 ‘To say that Boldrini is the same as Mastella, as Bossi’s son or as all those *people having to do with Silvio Berlusconi (including Berlusconi himself)/Berlusconi and Co.* is ideological violence that does not lead anywhere.’
- b. Meglio accozzaglia che *renzaglia*. (Magni 2018, 216)
 ‘Better being a jumble than a *renzaglia/a group like Matteo Renzi and his associates.*’

These productive ad hoc categories as well as findings on teenager’s language use and nonce-formations like It. *parentame* ‘kinship’ (cf. Poletto/Penello 2006) suggest that collection nouns as a derivational category are not as far-fetched a notion as seems at first sight. I will consequently present a new empirical approach focusing on these ad hoc derivations and basing my analysis on the theory of construction morphology. I will first give an overview on this theoretical approach, elaborating its central assumptions and justifying my choice of this theoretical background, and will then describe the methodology and results of my empirical analysis. The central research question for this morphological examination is the following:

RQ_{morph}: To what extent does the realised productivity of collective derivational processes reflect their expanding productivity shown in the number of newly coined words based on them? To what degree can the category of collection nouns then be defined as a sharply delimited derivational category?

5.1 Theoretical preliminaries: Construction morphology

In contemporary linguistics, there are numerous approaches, frameworks and theories on which one can rely when dealing with word-formation processes. In what follows, construction grammar (henceforth CxG), or more specifically construction morphology (henceforth CxM) will be chosen as a theoretical framework for the following analysis. I will first give an overview of this morphological approach and will then substantiate the choice for it.

The term *construction grammar* was coined by Fillmore/Kay/O’Connor (1988) and, in the years following this key publication, a number of approaches was developed that view language as a network consisting of form-function pairings. The following approaches are generally classified as constructional: the

Berkeley Construction Grammar (James Fillmore and colleagues), the *Sign-based Construction Grammar* (Hans Boas and colleagues), the *Radical Construction Grammar* (William Croft) and the *Cognitive Construction Grammar* (George Lakoff, Adele Goldberg) (cf. i.a. Traugott/Trousdale 2013, 2–8 for an overview). These approaches generally differ in whether they adopt a more formalistic or rather cognitive perspective on language, like the Berkeley Construction Grammar vs. the Cognitive Construction Grammar. Nevertheless, they share a number of similarities which will be focused on here. In this respect, Goldberg (2013, 15–16) lists the following:

- a) Human language consists of *constructions*, that is to say, conventionalised pairings of form and function that represent different degrees of abstraction. A derivational pattern like Engl. [V-*ing*] e.g. can be equally seen as a construction as well as the much more abstract and complex pattern of expressing the passive voice (cf. Croft 2001, 17; Goldberg 2006, 5).
- b) A constructionist approach concentrates on the surface structure of a construction and its constituting schemata. In this respect a construction like *I coughed a moth out* can be explained by the fact that although *to cough* is not a transitive verb, it follows the surface structure of such in being used as one (cf. i.a. Goldberg 2002).
- c) Constructions are not stored chaotically or as a list in the mental lexicon but are instead interconnected via a hierarchical network. This represents the different degrees of abstraction or schematicity of constructions and within which more specific constructions take on their characteristics by default inheritance from the respective superordinate constructions (e.g. the word order in prepositional phrases in English) (cf. i.a. Langacker 1987a; Goldberg 1995; Booij 2010).
- d) Some linguistic functions are universal because they serve cognitive and communicative needs (e.g. to describe events). How these functions are expressed is determined by the particular language and its specific inventory of constructions. As a consequence, these functional approaches concentrate on describing and comparing different means of expression – language universals are rather to be found external to grammar in this view (cf. i.a. Croft 2001; Haspelmath 2008).
- e) Speakers perceive utterances, words and phrases and from these they generalise patterns which can again be used to create new utterances, words and phrases based on them. It follows from this that constructions which are more frequent can be accessed more quickly and easily, that they are more immune to regularisation processes (cf. highly frequent irregular verb forms, for instance) and they can serve as a centre of attraction for analogy (cf. i.a. Tomasello 2003; Bybee 2010).

Especially the first aspect of this list of constructionist principles underlines the assumption that lexicon and grammar are not two distinct, clearly delimited linguistic domains, but should rather be understood as a continuum. The constructionist understanding of language is thus based on the premise of an interdependent network of linguistic components, as assumed e.g. in the model of parallel architecture by Jackendoff (2002) (cf. also Croft 2001, 14–15; Jackendoff 2013). In this model, the components of language – phonetics, syntax and concepts – are not only linked to each other via interfaces that follow specific rules but are also linked to extra-linguistic components such as hearing and vocalisation, perception and action. Following from this, it is improbable that language consists of independent building blocks, but we ought rather to assume that “all knowledge of language is encoded in terms of stored pieces of structure organized into an inheritance hierarchy, without a strict separation between lexicon and grammar” (Jackendoff 2008, 27). Authors, however, differ in terms of the particular components they count as constitutive in language. In the model of parallel architecture e.g. morphology is not represented as a component on its own, because it is seen as the “extension of the parallel architecture below the word level”: the morphological form interacts equally with all the other linguistic components, relating them to each other (Culicover/Jackendoff 2005, 19; cf. also Jackendoff/Audring 2020). William Croft, founder of Radical Construction Grammar, in contrast assumes morphology as well as pragmatics to be components in their own right (cf. Croft 2001, 18). The various aspects of CxG can be summarised as follows:

In C[onstruction] G[rammar], the grammar represents an inventory of form-meaning-function complexes, in which words are distinguished from grammatical constructions only with regard to their internal complexity. The inventory of constructions is not unstructured; it is more like a map than a shopping list. Elements in this inventory are related through inheritance hierarchies, containing more and less general patterns (Michaelis/Lambrecht 1996, 216).

Before coming more specifically to the understanding of morphology in the framework of CxG, a short note on terminology is needed. In constructionist literature, the terms *schema*, *construction* and *construct* are not necessarily used in a uniform manner. As indicated above, most authors agree on the syntax-lexicon continuum and on the existence of different types of construction, which, following the classification of Traugott/Trousdale (2013, 13–20), differ in their schematicity, productivity and compositionality. English passive constructions e.g. are highly schematic and determine variables which can be filled with specific lexical material, whereas a complex lexeme like *daredevil* is less schematic and cannot therefore license any other subordinate constructions. Goldberg (2006, 5)

assumes in this respect: “all levels of grammatical analysis involve constructions: learned pairings of form with semantic or discourse function, including morphemes or words, idioms, partially lexically filled and fully general phrasal patterns.”

Except for these general agreements, authors differ in how many levels of schematicity they propose, how they name them or whether they genuinely distinguish different levels at all. Traugott/Trousdale (2013, 16–17) e.g. differentiate *schemata*, *subschemata* and *micro-constructions*, Croft (2001, 26) only delimits schematic constructions from (more specific) constructions, while Goldberg (1995; 2006) makes no terminological distinctions between different levels of abstraction. Actual, attested occurrences of a schematic construction, i.e. what inheres in the relation between type and token, are either called *construct* (cf. Traugott/Trousdale 2013, 14–15) or *construction* (cf. Goldberg 1995; 2006; Croft 2001). In the midst of this terminological incoherence is the basic question of how specific a construction has to be to still qualify as a construction – in other words –, whether a lexeme like *daredevil* is to be viewed as a construction, or whether this would only apply to constructions which are schematic in opening blank spaces to be filled with particular lexical material. In the framework of CxM e.g. Booij always assumes constructions to be schematic and that the instantiation of such an abstract schema in a complex word should always be called a construct. Traugott/Trousdale (2013) however propose to also call lexemes (micro-) constructions, because these also represent conventionalised pairings of form and function.

In the framework of this present study, I adopt the following terminological conventions: according to the constructionist view of language, I call every conventionalised pairing of form and function a construction. These constructions show differing degrees of complexity and schematicity. In this respect, an abstract quantifier schema has to be classified as a construction, but also the particular lexeme *a bit* (cf. Traugott/Trousdale 2013, 7). Specific instantiations of constructions are called constructs. These can either be represented by token, i.e. instantiations of lesser schematic micro-constructions, but also by neologisms and nonce-formations, i.e. instantiations of more abstract constructions.

How can this general understanding of language be applied to derivational processes? In the framework of CxG, this question is addressed by the theory of CxM, established in large part by Booij (2010). Like other constructionists, he assumes that complex words are conventionalised pairings of form and function. These are based on superordinate derivational patterns or schemata, which not only express generalisations of already lexicalised words, but can also function as a blueprint for coining new words. Particular constructions and more abstract schemata are mentally stored via hierarchical links in a complex interconnected

2014, 76–77). This transcending of the boundaries of a constructional schema can thus lead to language change (cf. Traugott/Trousdale 2013).

The theoretical approach of CxG in general and of CxM in particular can consequently explain a variety of phenomena such as language acquisition (cf. e.g. Tomasello 2003), language change (cf. e.g. Traugott/Trousdale 2013) and also phenomena of language contact (cf. e.g. Boas/Höder 2018), because it equally takes into account language-specific patterns as well as universal aspects of thinking and speaking, like general cognitive principles of analogy and pattern recognition (cf. e.g. Marques de Sá 2001 for a comprehensive overview). In this framework, word-formation processes can be understood as representations of abstract constructional schemata, which are part of some wider constructional network. The intertwining of the different constituting components can occur not only via semantic but also via formal aspects (Langacker 2008, 226). In comparison to other morphological approaches, e.g. traditional generative morphology (cf. i.a. Scalise 1986), CxM thus takes into consideration not only structural and formal aspects of word formation, but focuses also on functional and semantic aspects. In this respect, CxM allows explanations of word-formation processes from an onomasiological perspective and enables me to answer my research questions, viz. which suffixes form collection nouns in Romance languages and to what degree this word-formation process is a productive one. This functional approach to the schematic representation of collection nouns will now be explained in more detail.

As already discussed, many Romance suffixes deriving collection nouns are highly polyfunctional. Zwanenburg (2000) explains this with recourse to French. Here, nouns designating states, qualities, processes and collections can be derived through various suffixes that in turn often assume more than one function. Fr. *-erie* e.g. derives denominal state nouns (*clownerie* ‘clownery’), deadjectival quality nouns (*brusquerie* ‘brusqueness’), deverbal process nouns (*badinerie* ‘badinage’) as well as denominal collective nouns (*verrerie* ‘glassware’). At the same time, deverbal process nouns can also be derived by means of *-ment*, *-at* and *-ure*, as in *aboisement* ‘barking’, *assassinat* ‘assassination’ or *ouverture* ‘opening’ (cf. Zwanenburg 2000, 843; cf. also Table 3.1). The interplay of synonymy and polysemy consequently forms a systematic network of form-function relations, which is governed and restricted by a number of different aspects. For the purpose of this present work, two issues are of primary importance. Firstly, there are effects of prototypicality, which determine one central function of a suffix. In this sense, other related functions are mostly metonymic or metaphorical extensions of this central meaning nuance. Booij (1986) illustrates this phenomenon with the example of Dutch agent nouns being derived on *-er*: the suffix prototypically derives *personal agent nouns* that denote a human agent like

Dutch *zender* ‘person who sends’, but may also derive *impersonal agent nouns* like *zender* ‘radio/tv station’ and instruments like *zender* ‘transmitter’ (cf. Booij 1986, 509–510). As has been described in chap. 3.2, this also applies to e.g. French collection nouns based on *-ment* which represent metonymic extensions from the prototypical meaning of events and processes. Secondly, the network of various polyfunctional suffixes is also governed by different degrees of productivity. As outlined before, productivity may either be understood as past productivity and thus a certain number of already lexicalised derivatives, or future productivity and thus a certain number of possible nonce-formations. Against this background, different suffixes may be represented with various degrees of productivity and frequency in the systematic network. For instance, as described in chap. 3.2, the Spanish suffix *-ar* is mostly restricted to locations of plants, so the past productivity is already (semantically) restricted, and because, as postulated by Rainer (1993, 409), the suffix does not form any neologisms, there is consequently no potential for future productivity either. Although the suffix may thus form collection nouns *lato sensu* like Sp. *manzanar* ‘collection of apple trees’ and is formally and functionally related to the suffix *-al*, it is not well represented in the network. With this in mind, an initial simplified constructional representation of COLLECTIVITY in French, based on the overview of Zwanenburg (2000, 843), may be sketched:

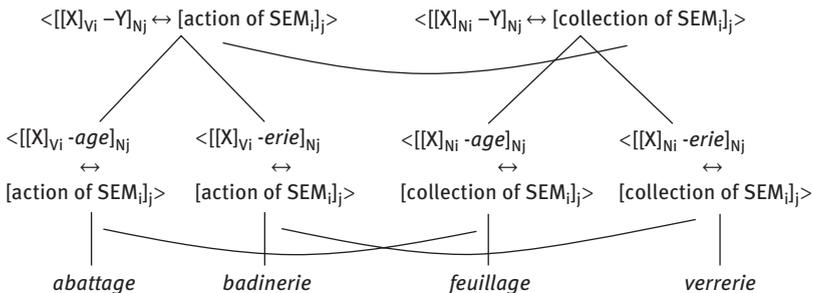


Figure 5.2: First draft of a constructional schema <collection> in French.

As shown in Figure 5.2, I assume a constructional schema <collection> which in this simplified example is concretised by the two sub-schemata $[[X]_{Ni} -age]_{Nj}$ and $[[X]_{Ni} -erie]_{Nj}$, which in turn can be instantiated by the two micro-constructions *feuillage* ‘foliage’ and *verrerie* ‘glassware’. Through metonymic links on their functional side, these collective sub-schemata are connected to another group of sub-schemata expressing the function of an action, which can be instantiated by the same suffixes *-age* and *-erie* and the micro-constructions *abattage* ‘felling’ and *badinerie* ‘badinage’. Figure 5.2 only takes into consideration two

possible constructional (sub-)schemata of the derivation of collection nouns in French and also only one possible type of link between different sub-schemata. It follows from this that the limits of this representational method of CxM are quickly reached and that it is not suited to representing the whole complexity of collection nouns as a possible derivational category not only in French, but in Romance languages in general. Moreover, it cannot account well either for prototypicality effects or for differing degrees of productivity (cf. *supra*).

Booij himself never explicitly depicted these semantic connections in his morphological networks. Although he discusses the systematic polysemy of e.g. Dutch agentive *-er* suffixes and problematises the different explanatory approaches, his constructional schemata always only contain one meaning component (cf. e.g. Booij 2010, 80). Rainer (2016, 348) also addresses this issue of “the absence of a direct expression of semantic relationship between schemas” as a potential problematic aspect in CxM. Consequently, I will adopt the theoretical implications of CxM in the following, but I have chosen to use a *semantic map* as a graphical representation to account fully for the network’s character and contours as well as the problematic issues addressed so far (cf. Luján 2010 for a similar approach on agent and instrument nouns). The concept of a semantic map was promoted mainly by Haspelmath (2003) to capture from a typological point of view the polyfunctionality of grammatical elements as well as language change and cross-linguistic variation. He exemplifies how such a map works by means of the functional scope of Engl. *to* and Fr. *à*. Both prepositions can mark directions, recipients and experiencers, Fr. *à* furthermore has the function of marking a predicative possessor (*ce chien est à moi* ‘this dog is mine’), and Engl. *to* additionally expresses purpose like in *I left the party early to get home in time* (cf. Haspelmath 2003, 213–215). The basic principles of such a map are the following: a function only appears on the map when there is at least one pair of languages that differ in how they use that function. The arrangement of functions in the map is governed by the polyfunctionality of the linguistic elements it represents. The related functions of a linguistic element have to be represented in a contiguous area on the map (cf. Haspelmath 2003, 217). Additionally, semantic maps are neutral to prototypicality effects, but theoretically can also represent them (cf. Haspelmath 2003, 232) and they may illustrate diachronic change, because one function or change often presupposes others and extension of functions often occurs directionally (cf. Haspelmath 2003, 232–237). Figure 5.3 illustrates these principles by means of the examples of Engl. *to* and Fr. *à*:

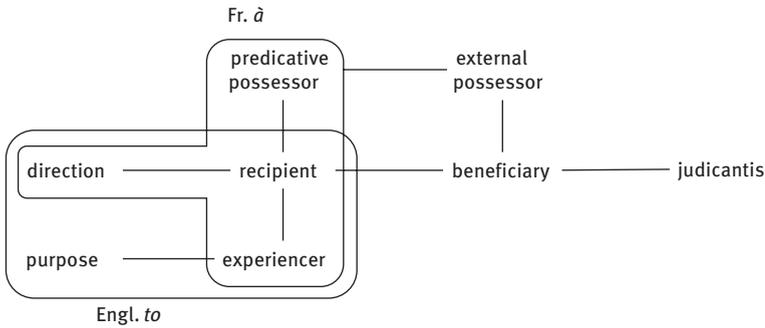


Figure 5.3: Semantic map of typical dative functions/functional scope of Engl. *to* and Fr. *à* (adapt. from Haspelmath 2003, 213–215).

Haspelmath himself is not interested in derivational processes, but only considers grammatical morphemes in his model of semantic maps. Following Luján (2010), however, there are many parallels between those issues related to the poly-functionality of grammemes and the derivational processes focused on by Luján and examined in this present framework. As exemplified by way of the derivational means of expression of agents, instruments and functions related to it in various languages, Luján (2010) shows that the functional ranges of derivational morphemes display a systematicity similar to the grammatical morphemes focused on by Haspelmath (2003). In this respect, he assumes, basing his analysis on languages like Basque, French, Greek or Turkish, a systematic polysemy between e.g. an agent-instrument function as well as diachronic paths of change going e.g. from the instrument to an additional means function. Some functions may only be displayed by a certain derivational morpheme in one language but not in another, while functional relations and paths of evolution are represented by contiguity in this map of derivational morphemes. The use of this method is justified by Luján (2010, 163) in its assumption of a continuum between lexicon and grammar. In this respect, “lexical and grammatical morphemes constitute a continuum, and their meanings are organized in the same way – inside a cognitive frame, we can assume that in both cases there are core and peripheral meanings, but that the borders between these meanings are synchronically blurry, which allows for transitions and semantic changes over time” (Luján 2010, 163).

Especially this last aspect illustrates that CxM and semantic maps are perfectly compatible, but that both also confer some advantage the other approach may not (essentially) have considered. An important advantage of semantic maps in comparison to the formulas of CxM is their ease of representation. A semantic map can summarise a high number of various functions and aspects to fully account for the semantic range of a grammatical or derivational morpheme in a

given set of languages. As we saw in Figure 5.2, this kind of complexity is not portrayable with the formula-based representation of CxM. In contrast, the kind of abstraction a semantic map represents leaves open many aspects relevant for my analysis. The schematic hierarchisation of CxG may also be depicted in a semantic map by the different contiguous areas. In this respect, an abstract schema <recipient> may be instantiated by the two more concrete sub-schemata of Fr. *à* and Engl. *to* in Figure 5.3. In opposition to this, semantic maps are not able – and are not designed – to display productivity and prototypicality effects, and neither do they show formational restrictions. That is to say, a semantic map shows all the functions that a morpheme, be it grammatical or derivational, can possibly fulfil, irrespective of its status in language use. Here CxG and its general assumptions come into play – which are not incompatible with semantic maps, but simply are not considered since other aspects are of greater importance (cf. Haspelmath 2003, 232). Most relevant in this respect is point e) of the aspects shared by all constructionist approaches described above. Given the usage-based approach of CxG, I also give great importance to the role of frequency and productivity: it is not only important to know what kind of functions a linguistic element may fulfil, but also to what extent. I have already shown that some functions of a derivational suffix are more central in either being more frequent or diachronically the original function. This is the case with e.g. Fr. *-ment* which originally and most frequently derives action nouns, but also has a more recent and less frequent secondary collective function (cf. chap. 3.2). Semantic maps can only express possible paths of evolution in assuming that “some changes presuppose others” (Haspelmath 2003, 233), but not prototypicality based on frequency. Taking into consideration the different methods of measuring productivity, frequency in this respect may refer both to frequency of hapax legomena and frequency of types and tokens.

After these considerations, sketching a semantic map of COLLECTIVITY in Romance languages, based on the lexicographic data summarised in Table 3.1, would be theoretically possible. But, as indicated above, this would only represent past functional extensions and past productivity. For this present approach focusing on present-day language use by analysing nonce-formations, the proposal for such a map will be presented after the following elaborations in chaps. 5.2 and 5.3.

Summarising the understanding of morphology adopted in this framework, (complex) words are understood as constructions – that is, conventionalised pairings of form and function. These constructions form a hierarchical network in which characteristics of more abstract constructions are transferred via default inheritance to more concrete constructions. These abstract constructions, or constructional schemata, allow speakers to generalise by way of a certain type of

construction, to therefore interpret it and to build new constructions on its base. In accordance with the usage-based principle of CxG, more frequent constructions are faster to access and to learn in language acquisition. According to the syntax-lexicon continuum, a construction needs not only be a syntactic means of expression but can also comprise simple words or lexicalised syntagmata. In the following, I will focus on complex words as the typical morphological type of collection nouns (cf. chap. 3.2). The interconnected network of constructional schemata can be represented by a semantic map because this can display continuous relations as well as hierarchies between various levels of abstraction. Such a map, however, may only be understood in addition to aspects not displayed, such as prototypicality and productivity.

5.2 Methodology

Having discussed the state of the art on derivational processes in collection nouns in chap. 3.2, as well as the understanding of morphology adopted here in chap. 5.1, I will now focus on my empirical study. The central issue I want to address in this analysis is that research on lexicalised collection nouns in general considers collection nouns not to represent a very productive word-formation process in Romance languages, either because collection nouns are only derived through highly polyfunctional suffixes or because the derivational processes are said to be no longer productive. As indicated above, research on nonce-formations in Italian, however, suggests that this is not necessarily the case.

In the following, I will present an empirical study on nonce-formations derived by means of the collective suffixes in Table 3.1. As in chap. 3.2, I will focus on French, Spanish and Italian. Taking into consideration the theoretical framework of CxM, there is the necessary assumption that these ad hoc derivatives allow conclusions about the understanding of the speakers of the single potentially collective suffix. Moreover, as already elaborated above, such an analysis allows for a more realistic assessment of the productivity of certain suffixes than a mere consultation of lexicographic information would achieve. Bauer (2001, 48–49) captures this aspect accordingly: “Type frequency is the result of past productivity rather than an indication of present productivity.” Hilpert (2014, 82), in turn, argues from this “that a low ratio of hapaxes indicate[s] the absence of productivity.” Whereas the consideration of lexicographic information in chap. 3.2 thus enables me to draw some initial conclusions and to indicate some first tendencies, only a sophisticated analysis of nonce-formations can help to paint a much more detailed picture of synchronic usage and productivity of the single derivational processes already discussed. Taken together, the two approaches can com-

plement each other. Finally, the results of such an analysis could additionally support a possible unidirectional lexicalisation path of collection nouns, an issue I will discuss in part III.

The empirical basis for this examination of nonce-formations the *TenTen* corpus family already introduced in chap. 4.1. The same advantages of these corpora mentioned in relation to the examination of OMNs also apply to this present morphological analysis (comparable token quantities, spontaneous speech data etc.). With respect to this special type of data a comment on the web-crawling *Spiderling* tool is, however, necessary. This tool is an algorithm that crawls the web and automatically discards everything a linguist might not want to analyse (e.g. URLs, lists etc.). As a consequence, the *TenTen* corpora comprise a huge amount of data⁵² which in general is mostly suited for quantitative analyses. For a qualitative analysis of ad hoc derivatives such as this present one, this sometimes leads to the problem that the corpus contains words that may be hapaxes, because they only occur once in the corpus, but which are not necessarily unique as nonce-formations. In the *itTenTen16* corpus e.g. I found the Italian word *cinesaglia* ‘chinaware’ which for the same period of 2016 can also be found in Google with various occurrences. *Cinesaglia* consequently is a hapax legomenon in *itTenTen16*, but it is not a unique ad hoc construct because other speakers seem to use it as well. The interpretation of the data therefore has to be carried out very cautiously. Like any other corpus analysis, the results only reflect the data contained within that specific corpus. Cases like *cinesaglia*, which were unique in the *itTenTen16* corpus have been treated as nonce-formations and therefore taken into consideration for the analysis.

For every suffix in Table 3.1, I created in each of the language-specific sub-corpora a list of unique words, i.e. hapax legomena, that end with these suffixes and which are tagged as nouns; this kind of search should find nonce-formations created only once for a special communicative purpose. To be able to handle the number of results, I limited the search to the singular.⁵³ An example of one possible configuration is found in Figure 5.4, illustrated by French *-ment*.

⁵² Exact number of words for each of the three corpora: *frTenTen12*: 9,889,689,889; *spTenTen11*: 9,497,213,009 and *itTenTen16*: 4,989,729,171.

⁵³ By creating a word list in the *itTenTen16* corpus, one can additionally differentiate between upper and lower case to avoid doublets. This feature is not available in either the *frTenTen12* or the *esTenTen11* corpus, so I also excluded this possibility for Italian to increase comparability. I furthermore disabled for every analysis the feature *include nonwords*.

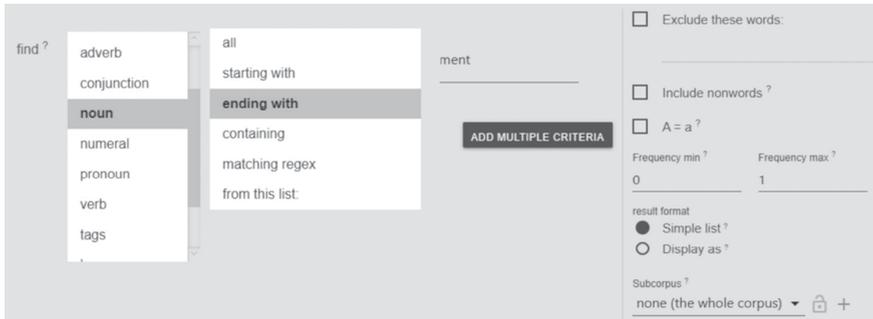


Figure 5.4: Example of the corpus configurations in *frTenTen12* for the suffix *-ment*.

The result for each suffix is a list of up to 1,000 words (the number of words that *Sketchengine* allows to download), this is the raw data for the analysis.⁵⁴ Yet many of the occurrences found in these word lists do indeed represent hapax legomena, but no ad hoc derivatives, in the sense of ‘new coined word’. This kind of hapax comprises examples of typing mistakes (cf. (54a): *hertamienta* instead of Sp. *herramienta* ‘tool’), graphically linked words (e.g. in URLs or because of a missing blank space; cf. (54b): *tuttovantaggio* instead of It. *tutto vantaggio* ‘every advantage’), the phonetic representation of mostly English words (cf. (54c)), or a superlative or negation of a lexicalised word (cf. (54d)). These occurrences were all eliminated. I also excluded all those cases like in (54e), where there is apparently some nonce-formation, but it is not interpretable because of a lack of context. In this exact example, it is not clear what *cette topshopperie* exactly refers to: it may be based on the proper name of the British fashion retailer *Topshop*, but apart from this, the context does not give me enough pieces of information to decide whether the nonce-formation refers to an action, a collection or something entirely different.

- (54) a. PERO bajo mi punto de vista, nuestro principal problema a dia de hoy es tener una *HERTAMIENTA FIABLE* con la que poder trabajar, [. . .] (*esTenTen11*)⁵⁵
 ‘But in my opinion, our central problem today is to have a reliable *tool* with which one can work, [. . .]’

⁵⁴ The data for this empirical study can be found here:

French: <https://hdl.handle.net/11022/0000-0007-EC55-0>

Spanish: <https://hdl.handle.net/11022/0000-0007-EC58-0>

Italian: <https://hdl.handle.net/11022/0000-0007-EC57-E>.

⁵⁵ In what follows, I will only indicate the respective corpus to cite the source, because the original websites are either not available anymore or generally not available for external users – the latter is often the case with forums having restricted access.

- b. rendendole applicabili al settore dell'acconciatura a *tuttovantaggio* del business salone. (*itTenTen16*)
 'making them applicable to the hairdressing sector to *the full benefit* of the salon business.'
- c. Mais aujourd'hui, cela s'appelle Mombay (prononcer *môme-baille*). (*frTenTen12*)
 'But today, it's called Mombay (do pronounce it *môme-baille*).'
- d. Algunos optimistas, si es que quedan unos pocos en relación a esta *megaingeniería* arquitectónica y de corrupción, dicen que el hito podría alcanzarse en 2010. (*esTenTen11*)
 'Some optimists, if only a few of them remain in relation to this architectural *mega-engineering* and corruption, say that the milestone could be reached in 2010.'
- e. Margaux@ oui, oui et un grand oui pour cette *topshopperie* et bien la naf naf n'est pas mal mais elle ne pourra pas éclipser mon sorbet lemon;) (*frTenTen12*)
 '@Margaux, yes, yes and a big yes to this *topshopperie* and the piece of clothing of Naf Naf [= French ready-to-wear brand] isn't bad, but it can't overshadow my lemon sorbet.'

All these occurrences not relevant for the present research question have therefore been eliminated, and the remaining occurrences of transparent ad hoc derivatives are captured and analysed. For each occurrence, I categorised the derivational base and the meaning of the derivative (according to the categories already established in Table 3.1). For additional analyses, I also examined the following issues: as elaborated in chap. 3.2, many collective nouns have a pejorative connotation and many of these negatively connoted collections are comprised of human referents. For this reason, I also classified the type of referent (e.g. human, animal, artefact), whether it has a pejorative connotation and whether the derivational base also already has one. By means of this classification, I will analyse the impression that (human) collections are especially prone to pejoration. I will also identify the respective determiner of the collective ad hoc derivatives to examine possible suffix preferences for either CCNs or OMNs. In the following, I will first consider the general results for all three languages and will then analyse the French, Spanish and Italian results in more detail.⁵⁶

⁵⁶ See Kleineberg (accept.) for a condensed version of this chapter focussing mainly on the analysis on Spanish.

5.3 Results

Table 5.1 below shows three sets of frequencies: Firstly, there are the absolute frequencies of transparent nonce-formations in the list of 1,000 words (Freq./1,000). These absolute frequencies are the basis of semantic analysis of the suffixes, because they represent the actual occurrences classified. Secondly, there is the frequency of transparent nonce-formations extrapolated to all hapaxes (Freq./hpxs). As shown in Table 5.1, the suffixes studied not only diverge in their proportion of transparent ad hoc constructs in the list of 1,000 words, but also in the overall frequency of hapaxes. For instance, Fr. *-ure* and *-at* derive a similar number of transparent nonce-formations in the list of 1,000 words (366 and 365 respectively). In contrast, the overall frequency of hapaxes, i.e. considering both transparent ad hoc constructs and other hapaxes as e.g. typing mistakes, diverges much between the two suffixes: there are 28,139 hapaxes for *-ure* and 52,135 for *-at*, thus twice as many. For this reason, I have extrapolated all absolute frequencies of transparent ad hoc constructs to this entire frequency of all hapaxes to be able to compare the degree of productivity between suffixes within the same language. Thirdly, there is the frequency of the extrapolated frequencies relative to the quantity of words of the entire corpus (Freq./crps; cf. footnote 52). To be able to compare the productivity of particular suffixes not only within the same language, but also between languages, I calculated the hypothetical frequency of each derived nonce-formation relative to the specific corpus size (extrapolated to 10,000,000). This means that I took the actual frequency of nonce-formations in the list of 1,000 words for e.g. Fr. *-ure*, which is 13, extrapolated this frequency to the total of hapaxes in the corpus, which is 366, and then calculated its relative frequency in the *frTenTen11* corpus, which is 0.37. The latter frequencies will allow the comparison of suffixes and the ad hoc constructs derived with them between languages and corpora. Keeping in mind that the latter two kinds of frequency are hypothetical – because they are extrapolations – these comparisons have to be taken with care, but at least can they give an impression of the productivity of particular suffixes compared to other representatives in the same language and to similar suffixes in other languages. The fourth column shows the range of meaning of every suffix studied. These indications are the result of the semantic analysis of the nonce-formations found in the corpora and represent their main semantic nuances. The order of the semantic values represents the dominance of a certain domain of meaning. For instance, It. *-aggio* primarily derives action nouns (ev.) and secondarily collection nouns (coll.). These tendencies will be discussed in more detail in the respective analyses of the three languages in chaps. 5.3.1, 5.3.2 and 5.3.3. For reasons of convenience, I only consider those

Table 5.1: Quantitative overview of the ad hoc derivatives based on potentially collective suffixes in the *TenTen* corpora.

Latin	French			Spanish			Italian						
	Suffix	Freq./1,000 hpxs	Freq./sem. crps	Suffix	Freq./1,000 hpxs	Freq./sem. crps	Suffix	Freq./1,000 hpxs	Freq./sem. crps				
-MEN	/			-ambre	10	10/287	0.01	coll.	-ame	81	308/3,806	0.62	coll.
-TŪRA	-ure	13	366/28,139	0.37	div.				-amen	35	39/1,109	0.04	coll./augm.
-MENTUM	-ment	6	180/92,929	0.18	ev.				-dura	150	236/1,571	0.21	ev./prop.
-MENTA									-mienta	1	1/265	0.00	other
-ĀTUM	-at	7	365/52,135	0.37	div.				-menta	3	3/380	0.00	div.
-ĀTA	-ée	3	88/29,195	0.09	other				-ada	71	1,353/19,053	1.23	ev./coll.
	-ade	50	595/11,901	0.60	ev./coll.				-ado	22	546/24,821	0.50	div.
-ĀLIA	-aille	44	154/3,492	0.16	coll./pej.				-alla	2	5/2,272	0.00	coll.
-ĀTICUS	-age	110	5,241/47,643	5.30	ev.				-ajie	164	1,020/6,220	0.93	ev./coll.
-ĀRIUS	-erie	185	2,666/14,412	0.19	ev./coll./loc.				-eria	175	937/5,354	0.85	ev./coll./loc.
									-erio	339	339/535	0.31	coll./ev.
									-aria	8	42/5,304	0.04	div.
									-ario	29	350/12,074	0.32	coll./loc.
									-aglia	56	74/1,284	0.15	coll./pej.
									-aggio	115	632/5,495	1.27	ev./coll.
									-eria	99	648/6,543	1.30	loc./ev./coll./prop.

functions with a minimum proportion of 10%, and the suffixes Fr. *-erie*, Sp. *-ería* and It. *-eria* as well as *-tura* have more than two dominant functions, viz. e.g. eventive, collective and locative. Also for reasons of convenience, I will only analyse in detail those ad hoc constructs of suffixes with an absolute quantity greater than 30. This excludes not only the very marginal cases of e.g. Fr. *-ment*, but also borderline cases like Sp. *-ado* (22) and *-ario* (29). These latter suffixes are excluded from the examination since they not only derive a relatively low number of ad hoc derivatives in my sample, but also because those words built upon them represent a high degree of functional heterogeneity. For *-ado* I found collection nouns like *futbolariado* ‘collection of football fans’ (< *futbolero* ‘football fan’) and ambiguous cases like *ennegrillado* ‘emphasis’ (< **ennegrillar* ‘to mark in bold type’ < *negrilla* ‘bold type’), which oscillate between an eventive and resultative reading – the verb *ennegrillar* is not lexicalised either. I also found other cases like *embellotado* ‘~ frame’ which is based on *bellota* ‘acorn’. In the respective context, this nonce-formation is presented as an option for *bastidor de madera* ‘wooden frame’,⁵⁷ one may thus imagine some kind of resultative reading of the constructed verb *embellotar* ‘to build out of acorns’. The difference between suffixes like Sp. *-ado* and *-ario*, on the one hand and It. *-eria*, which does not show any functional preference either, on the other, is that the former reflect a higher degree of complexity in their underlying formation processes. Here there are often intermediate steps (like the newly coined verb *ennegrillar*), whereas all the other nonce-formations in the corpora show a direct combination of a lexicalised base word and the respective suffix. Since pre-sorting the raw data is a time-consuming task, a separate analysis of Sp. *-ía* and *-ío* has not been possible.⁵⁸ In these cases, due to their homonymy to other suffixes (e.g. verbal *-ía* and *-ío*), the susceptibility to errors is so high that the quantity of irrelevant occurrences eclipses any analysable results. Furthermore, I did not consider Fr. *-ain*, *-in* and *-un* nor It. *-ato*, because these were already classified as only limitedly collectivising and/or unproductive (cf. Table 3.1). I classify every ad hoc construct as collective when it refers to a plurality of referents, which is the minimalist definition of a collection noun (cf. chap. 3.1).

57 “La colocación de los cristales en el embellotado o bastidor de madera distinguió a los vitrales criollos de sus similares foráneos, de ahí que muchos estudiosos lo incluyan como detalles representativos del escenario nacional.” (‘The placement of the glass sheets in the *embellotado* or wooden frame distinguished the Creole glass windows from their foreign counterparts, hence many scholars include it as representative details of the national scene’) (*esTenTen11*).

58 I sincerely want to thank María José Gallucci who helped me with the pre-sorting of the Spanish word lists.

Comparing the mere frequencies of nonce-formations in the three languages studied here, some general tendencies become apparent. The Sp. suffixes *-mienta* and *-menta* do not derive a considerable number of transparent ad hoc constructs in the sample analysed (rel. freq.: 0.00), whereas It. *-mento* shows a higher degree of productivity with a relative frequency of 2.31 ad hoc constructs in *itTenTen* – Fr. *-ment* lies between these two poles and has a mediocre degree of productivity (rel. freq.: 0.18). The expanding productivity thus mirrors the realised productivity of these suffixes as it has been discussed in the literature (cf. DES, s.v. *-miento*, *-mento*, *-menta* for the assessment of the productivity of the two suffixes in Spanish). In contrast, Sp. *-amen* only derives nonce-formations with a relative frequency of 0.04 in *esTenTen*. This contrasts with It. *-ame*, which derives ad hoc constructs with a relative frequency of 0.62 in *itTenten16*. Both suffixes are, however, said to be “productive” (cf. Poletto/Penello 2006; Rainer 2018b, 448), the expanding productivity reflected in the corpus data thus differs from the realised productivity described in the literature.

This brief overview of some general tendencies indicates that the degree of realised and expanding productivity of a particular suffix can either be identical – in this case, the corpus analysis mirrors the literature based mostly on lexicographic information – or different – in this case this present empirical analysis diverges from the tendencies indicated in the literature. Furthermore, suffixes that have the same Latin origin do not necessarily show the same degree of productivity or range of meaning when comparing them between languages. In the following, I will consequently focus on a mostly qualitative analysis of the nonce-formations in each language to thereby concentrate on cross-linguistic tendencies e.g. concerning pejorative connotations. The first domain of analysis will deal with the question of whether the potentially collective suffixes, at least in the Romance languages being examined here, can productively derive collection nouns or whether collection nouns are not in fact evidence of a productive derivative process in Romance language – as postulated until now.

5.3.1 French: Collective nonce-formations in *frTenTen12*

In the French sample, the four suffixes *-age*, *-erie*, *-ade* and *-aille* derive a sufficient number of transparent ad hoc derivatives to allow for a more detailed semantic analysis. The suffixes *-at*, *-ée*, *-ure* and *-ment* however derive hardly any transparent ad hoc derivatives in the sample (cf. Table 5.1). The few newly coined words based on these suffixes represent diverse meanings, mostly in accordance with the meanings of lexicalised derivatives as captured in Table 3.1. Examples are i.a. *boutonnure* ‘collection of buttons’ (< *bouton* ‘button’), *rassembleure* ‘the action of

gathering’ (< *rassembler* ‘to gather’), *désenbretonnement* ‘the rendering of so./sth. less Breton’ (**désenbretonner* ‘to make less Breton’ < *breton* ‘Breton’) or *pressurement* ‘the action of pressing’ (< *pressurer* ‘to press’). The semantic range of the four *productive* suffixes is depicted in Figure 5.5. I take the frequencies of nonce-formations extrapolated to all hapaxes as point of reference in order to compare the productivity of the different derivational patterns in French.

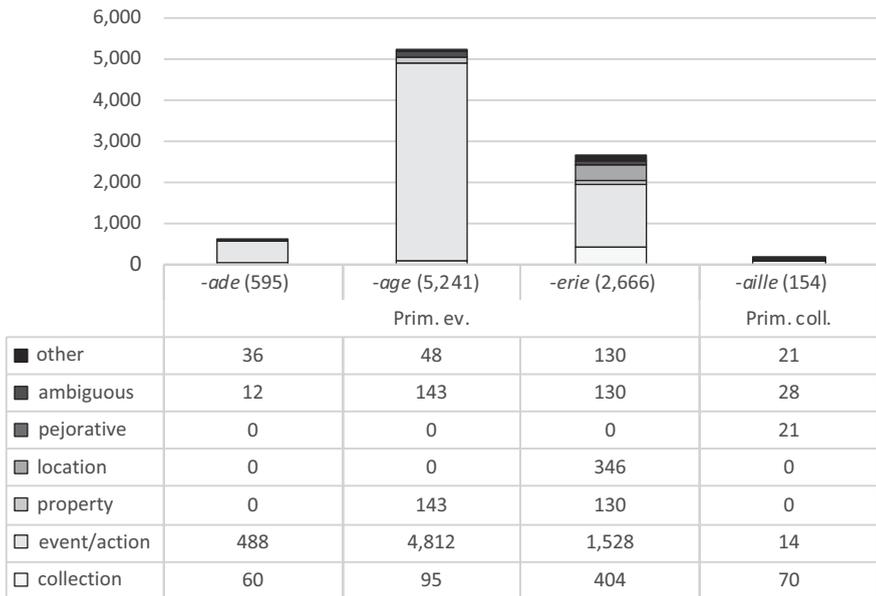


Figure 5.5: Semantics of unique nonce-formations in *frTenTen12* (1,000-word sample extrapolated to all hapaxes).

As already indicated in Table 5.1 and as shown in Figure 5.5, the derivatives on *-age*, *-erie* and *-ade* can be classified as primarily eventive and the derivatives on *-aille* as primarily collective. In what follows, I will analyse each suffix and the ad hoc derivatives based on it in turn.

Figure 5.5 clearly shows that the derivatives based on *-age* and *-ade* have nearly exclusively an eventive or process-orientated interpretation. The two suffixes differ in the fact that the nonce-formations on *-age* are mainly based on verbs (cf. (55a)), whereas *-ade* has a more heterogeneous group of derivative bases. Here, there is a number of derivatives based on proper nouns (cf. (55b)), common nouns, verbs and even other suffixes (*odontiade* < *-odontie* ‘-odont/related to teeth’). In contrast, there are not many collection nouns in the sample and the few cases one does find are rarely that clear, as in example (55c).

- (55) a. *Appellage* d'amis, Parfumage, Rasage, *Fringuage*, Fromage, Nicolas Cage et c'est parti pour sortir dans le bar branché du coin ! (*frTenTen12*)
'Calling of friends, perfuming, shaving, *dressing up*, cheese, Nicolas Cage and we are ready to leave for the trendy local bar!'
- b. Et il y avait bien sûr le fameux plan "B", la désormais célèbre *fabiusade* (*frTenTen12*)
'And there naturally was the famous plan "B", the henceforth well-known *fabiusade* (= 'the acting of Laurent Fabius'). . . '
- c. Moi par exemple, je me sens pleinement identifié à ma petite ville d'Asturies, son paysage: la mer, les montagnes (Picos d'Europe), sa gastronomie de produits naturels, et son "*paysannage*" (ses habitants). (*frTenTen12*)
'I, for example, feel fully identified with my small town of Asturias, its landscape: the sea, the mountains (*peaks of Europe*), its cuisine of natural products, and its "*paysannage*" (its inhabitants).'

In (55a), the two ad hoc derivatives of *appellage* 'calling' (< *appeler* 'to call') and *fringuage* 'dressing up' (< *fringuer* 'to dress up') represent morphologically analogous constructions to the lexicalised *rasage* 'shaving' and phonologically analogous formations to *fromage* 'cheese' and *Nicolas Cage* (supposedly [kaʒ] like [frɔmaʒ] and not [keʒdʒ]).⁵⁹ Both ad hoc derivatives are therefore typical deverbal nouns, reflecting the prototypical function of *-age*. It can be consequently confirmed what already has been assumed on the basis of lexicalised derivatives on *-age*, namely that "[. . .] jeder Ablt. auf *-age* die Vorstellung einer Handlung zugrunde[liegt]" ('the idea of an action underlies each derivate on *-age*') (Baldinger 1950, 42). Example (55b) equally illustrates the primarily process-orientated interpretations of derivatives on *-ade*. In this case, the derivate is based on the proper name *Laurent Fabius*, the former foreign minister of France, and expresses an action typical for him. It can therefore be concluded for the two suffixes *-ade* and *-age* that they synchronically only derive event nouns, and that they have no collectivising potential.

The group of ad hoc derivatives on *-erie* on the contrary is more heterogeneous. This suffix predominantly derives event nouns in about 57% of the sample (cf. (56a)). The other half is represented by nonce-formations with a collective interpretation (cf. (56b); 15%), locatives in the sense of '(work)shop' (cf. (56c); 13%), properties (cf. (56d); 5%) and a number of ambiguous ad hoc constructs that oscillate between these main groups (cf. (56e)).

⁵⁹ *Parfumage* equally does not appear in the dictionaries, but it occurs in *frTenTen12* in a total of 138 cases and was therefore not considered for this present analysis.

- (56) a. Cette douceuse *dansoterie* arcadienne est interrompue par l'arrivée de Faust et de Méphistophéla, qui entrent en scène dans leur costume de danse, [. . .] (*frTenTen12*)
 'This sweet Arcadian *dancing* is interrupted by the arrival of Faust and Méphistophéla, who enter the stage in their dance costume, [. . .]'
- b. [. . .] et rage encore face à cette *umpèterie* pancratienne, qui continue son oeuvre de maléfaction honteuse. . . (*frTenTen12*)
 '[. . .] and rage also before this almighty *UMP and its associates*, which continues its work of shameful malefaction. . .'
- c. Si tu à des magasin comme Soho (des *cadeauterie* je crois que ça s'appelle) dans les galerie Lafayette y'avais aussi des truks Hello Kitty, [. . .] (*frTenTen12*)
 'If you have shops like Soho (*gift-shops* I think it's called) in the Lafayette gallery there were also Hello Kitty-things, [. . .]'
- d. Sa révolte n'a pu prendre aucune forme. Allez savoir pourquoi! Paresse? Incompétence? *Foutraquerie*? Toujours est-il que le Sans Cause est juste un Sans Cause. (*frTenTen12*)
 'His revolt could not take any form. Who knows why! Laziness? Incompetence? *Craziness*? Anyway, the No Cause is just a No Cause.'
- e. Oui, nous leur devons une redevance de 1% de notre chiffre d'affaires pour les frais de publicité. Cette publicité prend en charge les affichages magasin, la *sachetterie*, les stickers pour les vitrines, les cartes de fidélité clients, les promos de mise en avant etc . . . (*frTenTen12*)
 'Yes, we owe them a fee of 1% of our turnover for advertising expenses. This advertising supports store displays, *bags/bagging up*, stickers for shop windows, customer loyalty cards, pre-promotions etc. . .'

As in the case of *-ade*, the eventive derivates on *-erie* do not necessarily have to be deverbal as in (56a), where the derivate is based on *dansoter* 'to dance only a little bit'. There are also ad hoc constructs based on proper nouns like *Socrate*, common nouns like *sous-marin* 'submarine' or even English bases like *to wipe*. The collective derivates mostly designate human collections like in (56b) where the base is the French conservative party UMP. Other collective derivates are e.g. *beaufrèrerie* 'collection of brothers-in-law' (< *beau-frère* 'brother-in-law') or *gaullatrerie* 'adherers of Charles de Gaulle and what he stands for' (< **gaullâtre* 'like Charles de Gaulle (pej.)' < Charles de Gaulle). An example of an inanimate collection noun is *make-uperie* which designates a collection of make-up. Locations like in (56c) are mostly based on common nouns and designate the things produced or sold there. In this sense, *cadeauterie* designates a gift-shop (< *cadeau* 'gift'), *jeanerie* a shop that sells jeans and *mascotterie* an atelier where mascots

are designed and made. Finally, property-denoting nonce-formations are mainly based on adjectives, as in *foutraque* ‘crazy’ in example (56d) or also common nouns like *cambrousse* ‘sticks/boondocks’ (> *cambrousserie* ‘being provincial’). Concluding this analysis of the suffix *-erie*, it can be stated that although it is highly polyfunctional both in lexicalised derivatives and in nonce-formations, it indeed has a collectivising potential, uniting the referents of the base noun in one collection.

As described in chap. 3.2, the suffix *-aille* is the only one that predominantly shows a collectivising function. The empirical analysis of ad hoc constructs in *frTenTen12* indeed reveals that *-aille* mainly derives collection nouns (45%), but that there is also a range of ambiguous cases. The collection nouns are mostly based on human-related common nouns and often, but not necessarily, have a pejorative connotation. There are equally pejorative collection nouns that already have a negatively connotated base noun (cf. (57a)) and those which have a neutral base (cf. (57b)). Note that *cambrousse* ‘sticks/boondocks’ appears twice as a derivational base in my corpus, once with *-erie* and denoting a property and once with *-aille* and denoting a collection. In chap. 5.3.4. I address in more detail this aspect of various nonce-formations that have same basis. In addition to these collective ad hoc constructs, some also occur where the suffix *-aille* only adds a negative connotation but does not pluralise the referents. This is illustrated in (57c), where the reference to a single person is obvious, and where a collective interpretation is not possible. Nevertheless, many of the nonce-formations on *-aille* cannot be unambiguously classified. In (57d) e.g. it is not clear whether *genouaille* (< *genou* ‘knee’) has to be interpreted as a collection noun, i.e. both knees (possibly with an additional pejorative sense), or whether the negative connotation predominates, where in this case there would be no collective interpretation. Put differently, it is not clear in this example whether the speaker bumped only one knee or both.

- (57) a. La *cambroussaille* vote socialvert. (*frTenTen12*)
 ‘The *totality of hillbillies* votes social-democratic-green.’
- b. en fait ce que l’*intellectuellaille frankaouie*, la plus médiocre de tous les pays démocratiques à part les scientifiques, modelée dans la réaction petite-bourgeoise, [. . .] ne supporte pas voir [. . .] (*frTenTen12*)
 ‘actually, what *the bunch of French intellectuals*, the poorest of all democratic countries except scientists, shaped by the petty-bourgeois reaction, [. . .] can’t bear to see [. . .]’
- c. Je crains pas ça tellment moi cette *moustiquaille* qu’on baptisa Raymond d’un père dit Queneau (*frTenTen12*)
 ‘I’m not so afraid, me, that *damn mosquito* which was named Raymond of a father who was called Queneau’

- d. points négatifs: je me suis pété la *genouaille* (*frTenTen12*)
 ‘negative points: I bumped my *damn knee/my knees*’

Summarising the analysis of French ad hoc constructs, the following can be stated: only four of the suffixes in Table 3.1 derive nonce-formations in a considerable quantity in the sample of *frTenTen12*. Of those four, three suffixes (*-age*, *-ade* and *-erie*) have a predominantly eventive function, with varying quantities of ad hoc constructs and different balancing of the semantic domains. The only suffix that predominantly derives collection nouns is *-aille*. This, however, is the least productive and has a rather high proportion of pejorative secondary meanings. What has already been assumed for the analysis of lexicalised derivates on these suffixes should be consequently confirmed for French: collection nouns are not a productive derivational pattern in French, neither in terms of realised nor of expanding productivity.

This corresponds to a similar phenomenon in French, namely the derivation of diminutives. In this respect, it has also been asserted that the synthetic derivation of diminutives in French is no longer productive and that there is instead the tendency to diminish analytically with aid of adjectives like *petit* ‘small’, *mini-* or *bébé* ‘baby’ (cf. Hasselrot 1957; 1972). The foundation for this tendency to become increasingly a language with an analytic linguistic architecture is said to have been already laid in Medieval French and the subsequent continuous erosion of inflexional and derivational suffixation like the plural *-s* (cf. Dębowski 2014, 134–143). By analogy, the gradual erosion also of French derivational suffixes may have led to a less synthetic and more analytic pattern for expressing the concept COLLECTIVITY and, in addition, a higher degree of polyfunctionality of those suffixes that are still productive. For instance, the descendants of Lat. *-MEN* are phonologically extremely reduced in French, e.g. *-ain* /*ɛ̃*/, and no longer productive (cf. Baldinger 1950, 79–82). This is not the case for their Spanish and Italian equivalents, as shown in Table 5.1. I will therefore point to the possibility that the lack of a collective derivational pattern in French may be ascribed to the general tendency of French to form analytic constructions rather than synthetic derivations. More research will need to be done to prove this hypothesis empirically.

5.3.2 Spanish: Collective nonce-formations in *esTenTen11*

As in the case of French, three main groups of ad hoc constructs in the Spanish corpus *esTenTen11* can be determined: first are those derivates that predominantly have an eventive interpretation, with this group represented by *-ada*, *-aje*, *-ería* and *-dura*. Second are those suffixes that mainly derive collective nonce-formations

(*-erío* and *-amen*). The third group consists of some suffixes that do not productively derive nonce-formations in the corpus and/or which do not show any clear semantic tendency (*-alla*, *-ambre*, *-menta*, *-mienta*, *-ada*, *-ario*, *-aria*) (cf. Table 5.1). The suffix *-ambre* e.g. also predominantly derives collective ad hoc constructs in the sample, like *hilambre* ‘collection of threads’ or *huesambre* ‘collection of bones = skeleton’, but the overall quantity of only 10 transparent ad hoc constructs is too small to make any generalisations about this suffix. The overall quantities of nonce-formations and the different semantic domains of the productive suffixes are shown in Figure 5.6. I will first describe the ad hoc constructs of the predominantly eventive suffixes and will then come to the mainly collectivising ones.

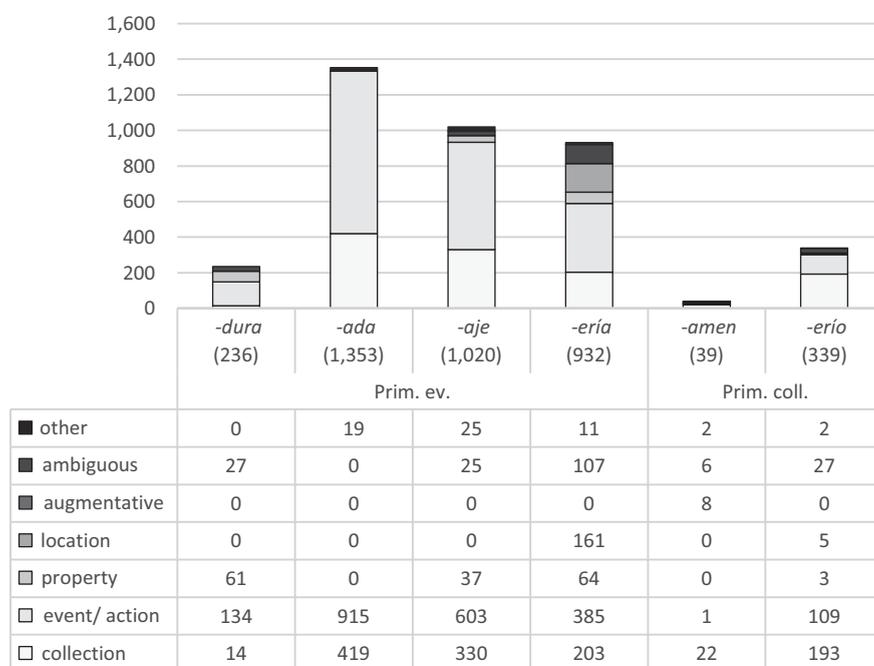


Figure 5.6: Semantics of unique nonce-formations in *esTenTen11* (1,000-word sample extrapolated to all hapaxes).

As shown in Figure 5.6, every primarily eventive suffix has a secondary collectivising function. The suffixes *-aje*, *-ería* and *-dura* additionally derive ad hoc constructs that designate properties or locations, while the ambiguous cases oscillate between these four main functions. As in French, eventive nonce-formations mainly differ in their derivation bases. Constructs based on Sp. *-ada* are predominantly formed upon proper nouns and consequently designate an action typical for this person.

In (58a) e.g. the ad hoc construct *bielseada* is based on the proper noun *Marcelo Bielsa* who coached the Argentinian national football team which was eliminated in the first round of the 2002 World Cup. Consequently, this nonce-formation is to be understood as ‘the catastrophe Marcelo Bielsa caused in 2002’. In contrast, eventive ad hoc constructs on *-dura* are mainly deverbal (cf. (58b), *engordura* < *engordar* ‘to increase weight’). This example also shows clearly that the ad hoc constructs do not necessarily have to be expressive, as was the case in many of the previous examples, but that they are sometimes simply created because of lexical gaps. Finally, nonce-formations on *-aje* or *-ería* have very diverse derivational bases, even syntagmata like *es mío* in (58c) or other event nouns as in (58d).

- (58) a. mirá si le dí tanta bola a lo que pasó después de la *bielseada* que no miré más el mundial salvo parte de la final y Brasil-Inglaterra, [. . .]
(*esTenTen11*)
‘we will see if I pay so much attention to what happened after *the catastrophe caused by Marcelo Bielsa* that I will not watch the World Cup except for the final and Brazil-England, [. . .]’
- b. a veces el ovario poliquístico evita que engordemos o a muchas mujeres le facilita la *engordura* (*esTenTen11*)
‘sometimes the polycystic ovary syndrome keeps us from getting fat or for many women it facilitates *weight gain*’
- c. Hace más de un año, en el Instituto Electoral Veracruzano, algunos empleados fueron sorprendidos con material que los involucraba en “*es-mionaje*” a compañeras del IEV. (*esTenTen11*)
‘more than a year ago, at the Veracruz Electoral Institute, some employees were surprised with material that involved them in actions of “*this-is-mine*” of feminine IEV colleagues.’
- d. Hay todo un maravilloso discurso de Morbius que – mas allá de su *blabletería* técnica – tiene suficientes puntos de contacto con conocimientos que el espectador posee [. . .] (*esTenTen11*)
‘there is a whole marvellous speech by Morbius that – beyond his technical *blabbering* – has enough points of contact with knowledge that the spectator possesses [. . .]’

The property-denoting ad hoc constructs on *-aje*, *-ería* and *-dura* are mostly not directly derived from property-denoting adjectives but are rather derived through a metonymic shift. *Empachadura* in (59a) e.g. goes back to the verb *empacharse* ‘to gorge oneself’ and consequently designates the state of having overeaten on something. Many of the hapaxes legomena in this semantic category are ambiguous between a property and another semantic interpretation, like e.g. a collec-

tion. In (59b) there are consequently two possible interpretations: the first is the property of vivacity and the second refers to people being vivacious – both being able to consume a society (this example is about people cheering and looking away, although there are issues like corruption etc.). Examples as clear as (59c) are consequently rare in my sample.

- (59) a. si que lo hago las papitas fritas los chicitos o cheestrees, tres platos como minimo y luego la sensacion de culpa o de *empachadura*. (*esTenTen11*)
 ‘yes, I have the French fries, small glasses of wine and the cheese puffs, three plates at minimum, and then the feeling of guilt or of *having over-eaten*.’
- b. Tampoco fuimos capaces de levantarnos cuando contemplamos cómo la *vivarachería* iba carcomiendo a una sociedad, hasta podrirla de abuso y corrupción. (*esTenTen11*)
 ‘we were also not capable of standing up when we complained about how the *vivacity/vivacious people* were bit by bit consuming a society, until decaying it with abuse and corruption.’
- c. Al término de su *adictaje* sirvió en diversas unidades hasta que como Comandante de Grupo, [. . .] (*esTenTen11*)
 ‘at the end of his *addiction* he served in various units even as commanding officer, [. . .]’

The collective ad hoc constructs derived by these predominantly eventive suffixes are mostly based on common or proper nouns indicating the ability to derive not spontaneous collection nouns but functional ones. As in the French examples, the denominal nonce-formations group the underlying referents of the base noun together (cf. (60a)), but the ad hoc constructs based on proper nouns take this named individual as the centre of attraction in the sense of ‘X and associates’ or ‘X and Co.’ (cf. *supra*). In (60b) e.g. the neologism *Fujimontesinada* is based on the proper names of Alberto Fujimori (president of Peru from 1990 to 2000) and Vladimiro Montesinos (chief of Peruvian secret service under him) and thus designates people who are associated with or are in the entourage of these two. The nonce-formation may thus not only imply a collection associated with this central attraction point but may also point to everything that has to do with them, that is corruption, bribery and crimes against humanity. *Caviarada* (< *caviar*) in this example is not found in the dictionaries either, though has not been considered in this analysis as it appears 326 times in *esTenTen11*. Only in the case of the suffix *-dura* are there deverbal collection nouns which then express the collective result of an action or people carrying one out. In this sense, *encabulladura* designates the result of *encabullar* ‘to unite

something with agave fibres', i.e. some cord made of agave fibres (cf. (60c)) and *linchadura* 'lynch mob' in (60d) goes back to *linchar* 'to lynch'.

- (60) a. Esos cartelones tan atildados contrastaron notoriamente con las telas raídas del *banderaje* de las organizaciones de izquierda que había acampado en la Avenida de Mayo un par de días antes. (*esTenTen11*)
 'Those classy posters contrasted notoriously with the shabby fabrics of the *flags* of the organisations of the left that had camped in the Avenida de Mayo a couple of days before.'
- b. Al menos soy eso y no un triste paria que un día se acuesta con la *Fujimontesinada* y otro amanece con la caviarada. (*esTenTen11*)
 'At least I am that and not a sad outcast who one day lies down next to the *bunch of adherents of Fujimori and Montesinos* and the other wakes up next to the left-wingers who don't want to give up their capitalistic behaviour.'
- c. Es el nombre con que se designa el moño de las cabulleras de la hamaca. En ellos va inserto el guaral de la *encabulladura* [. . .]; (*esTenTen11*)
 'It is the name with which one designates the knot of the agave fibres of the hammock. One inserts in it the cord of the *collection of agave fibres* [. . .]'
- d. El término hace referencia a una de las principales facciones dentro de las Fuerzas Armadas brasileñas, que se dividían básicamente entre los miembros de la "*linchadura*", [. . .] (*esTenTen11*)
 'The term makes references to one of the main factions of the Brazilian armed forces, which are divided basically into the members of the "*lynch mob*", [. . .]'

The only suffix that also derives locations is *-ería*. Like Fr. *-erie* it forms mostly derivatives for shops or workshops, and the derivational base consequently mostly represented by nouns designating things that are sold or made there. Examples of this type are *browneria* 'café which sells brownies' (< *brownie*), *mandillería* 'leather apron workshop' (< *mandil* 'leather apron') or *antigüería* 'second-hand bookshop' (< *antiguo* 'antique').

This section looking at the predominantly eventive ad hoc derivatives can be summarised as follows: in addition to their main function of expressing actions and events, the four Spanish suffixes *-ada*, *-aje*, *-ería* and *-dura* represent a similar division of labour between the semantic domains of collections, properties and to a certain extent locations. This group of Spanish suffixes thus behaves exactly like the French equivalents of *-ade*, *-age* and *-erie*.

The second group of suffixes has a predominantly collectivising function. In Spanish, this is the case with *-erío* and *-amen*. The first suffix *-erío* derives collective nonce-formations in 57% of cases in the sample from *esTenTen11*, and its secondary function is eventive (32%). The derivational base of the collective nonce-formations is nominal in the majority of cases and covers all semantic domains. (61a) e.g. represents a human collection of parents, (61b) of discrete inanimate referents (*tumba* ‘grave’ > *tumberío* ‘graveyard’) and (61c) represents a case where an abstract noun is collectivised (*mensaje* ‘message’ > *mensajerío* ‘messages’). The eventive ad hoc constructs on *-erío* are furthermore also based on various word categories, equally on common nouns (*basurerío* ‘talking rubbish’ < *basura* ‘trash’), verbs (*llamerío* ‘yelling’ < *llamar* ‘to yell’) and adjectives (*borracherío* ‘booze-up’ < *borracho* ‘drunk’). With the example of *abuelerío* in (61d) there is an analogous construct to *niñada* ‘childishness’ and thus expresses the typical behaviour of grandparents.

- (61) a. “Toy Story 3” continúa en el número 1 de nuestra taquilla por segunda semana consecutiva y metiéndose en el bolsillo a cualquier tipo de público, no importa la edad, chavalería o *padrerío*. (*esTenTen11*)
 “‘Toy Story 3’ continues to be number one in our box offices for the second week in a row and has won the favour of any kind of audience, no matter the age, young people or *parents*.’
- b. Con su hermano Lorenzo alguna vez se juraron no volver a pisar jamás el *tumberío*, pero su madre se los había pedido encarecidamente en su lecho de muerte: “El día que llegue el turno de Negrita, entiérrrenla conmigo.” (*esTenTen11*)
 ‘With her brother Lorenzo they once swore they would never set foot again in the *graveyard*, but her mother had urged them on her deathbed: “The day Negrita’s turn comes, bury her with me.”’
- c. confiaba que muchos no entendieseis lo de Jesus, tu gloria y demás *mensajerío* divino. . . . (*esTenTen11*)
 ‘I trusted that many of you would not understand Jesus, your glory and other divine *messages*. . . .’
- d. Y la mayor parte de las veces, a lo largo de las noches calurosas, después de la jornada de niñadas, *abuelerío* o, simplemente actividad diurna bajo el tórrido verano, [. . .] (*esTenTen11*)
 ‘And most of the time, throughout the hot nights, after the day of childishness, *behaving like grandparents* or simply daytime activity under the torrid summer, [. . .]’

As shown in Table 5.1 and Figure 5.6 above, *-erío* derives relatively many ad hoc constructs in the Spanish sample, especially compared to those based on *-amen*

and also compared to the relations between primarily eventive and collective suffixes in French (cf. Figure 5.5). The suffix seems to take over functions of other, unproductive Spanish derivational suffixes. In *esTenTen11*, there is e.g. the nonce-formation *roblerio* ‘oak forest’ (*roble* ‘oak’), which co-occurs with the lexicalised *matorral* ‘thicket’ and *pastizal* ‘pasture’. In this example *-erio* takes over the function of *-al* of collectivising plants (cf. *supra*) and consequently fills a gap, because it expands into a semantic niche that is no longer productive. This aspect confirms an assumption of David Pharies, who ascribes a high degree of productivity to *-erio*, not only in Latin America, but also in European Spanish (cf. DES, s.v. *-ío*; cf. also Ponce de León 2016). I will tackle the issue of diatopic variation after considering all the functional ranges of the suffixes studied here.

Although the suffix *-amen* also predominantly derives collection nouns (56%), the number of transparent ad hoc constructs in the corpus is much smaller than that of *-erio* (39 compared to 339, considering the extrapolated frequencies). As already indicated in Table 5.1, the Latin suffix *-MEN* has two possible continuations in Spanish. The popular suffix *-ambre* only derives 10 ad hoc constructs in *esTenTen11*, though mainly with a collective function (cf. *supra*). The cultism *-amen* however derives a considerable quantity of also predominantly collective nonce-formations – I can therefore empirically confirm what has been stated for lexicalised nouns on the continuations of Lat. *-MEN* (cf. Rainer 2018b, 448). The suffix *-amen* particularly shows functional collectivity, i.e. the derivational base is mainly nominal. Ontologically, there are mostly human referents as in (62a) (*lolitamen* ‘lolitas’) and discrete inanimate referents as in *cableramen* ‘tangle of cables/cable spaghetti’ (cf. (62b)). It is noteworthy that many nonce-formations on *-amen* in the corpus show the semantically logical extension to augmentatives, especially in the case of human body parts (cf. also Rainer 2018b, 448). This can be seen in (62c) where a collective interpretation of *bocamen* ‘big mouth’ (< *boca* ‘mouth’) is not possible, but also in (62b) we can think of an augmentative interpretation in the sense of ‘big cable’.

- (62) a. Un único apunte para que imaginéis cuan tormentosa debió ser mi adolescencia entre tanta *lolitamen* rubia y Tadzios viscontinianos fallidos [. . .] (*esTenTen11*)
 ‘A single note for you to imagine how stormy my adolescence must have been among so many blonde *lolitas* and failed Viscontinian Tadzios [. . .]’
- b. El computador y la tele en sendas mesas que pusimos pegaditas a la muralla de la casa, para poder sacar por la ventana el *cableramen*. (*esTenTen11*)
 ‘The computer and the TV on both tables, which we attach to the wall of the house, to be able to pull the *collection of cables/the big cable* through the window.’

- c. Bueno, Don Bob, que te parece el look de la “Courtney love”, sobre todo el *bocamen*, pues eso, felicidades a ti, Don Bob. (*esTenTen11*)
 ‘Well, Don Bob, how do you like the look of Courtney Love, especially the *big mouth*, well, congratulations to you, Don Bob.’

Summarising the Spanish data, in addition to the highly polyfunctional suffixes like *-ería*, Spanish makes use of a predominantly collectivising suffix, namely *-erío*. In contrast to French *-aille*, however, this may be considered as productive. Furthermore, there seems to be no preference for pejorative connotations in this suffix, a fact which was indeed obvious for Fr. *-aille*. Although *-erío* does show a primarily collective function, the relatively high proportion of eventive ad hoc constructs of about a third compared to two thirds has to be recognised (cf. Figure 5.6). It should be admitted finally that I did not, as mentioned above, clearly differentiate between *-erío* and *-ío* in the analysis in the first place. I searched for nouns ending on *-erío*, which thus automatically included nouns that end in *-ío*, like *pasajero* ‘collection of passengers’ based on *pasajero* ‘passenger’. A later analysis of the nonce-formations based on *-erío* reveals that indeed the majority of these transparent ad hoc constructs end in *-erío* and not in *-ío* (320 as against 20).⁶⁰ The overall analysis of the Spanish ad hoc derivatives in *esTenTen11* thus indicates that the concept of COLLECTION in fact is a derivational category in Spanish, as represented by the suffix *-erío*.

Before coming to the detailed analysis of Italian ad hoc constructs, a short note on diatopic variation in the Spanish sample is necessary. By examining the domain names of the source websites where the various nonce-formations were found, one central tendency can be determined: the great majority of the ad hoc constructs in general, collective or not, has its origin in Central and South American Spanish; only the suffix *-erío* is also present in European Spanish and the suffix *-amen* is mainly found in Europe and not the Americas. This finding has to be treated with caution, because the analysis of the respective domain names is no guarantee of the dialect of the individual speaker, but there does indeed seem to be the tendency for Latin American Spanish to be more creative and for *-amen* to be mostly restricted to European Spanish creative derivational processes. Such an analysis was not possible for French or Italian, because there is only variation between *.fr/.it* and *.com/.org*, so I cannot deduce any preferences for e.g. either European or Canadian French (the domain *.ca* was not present in my sample). Any kind of diatopic variation within France or Italy cannot be detected only with the aid of domain names.

⁶⁰ In the case of two ad hoc derivatives, I could not for certain clearly identify the derivational base, and consequently I could not determine if these ended in *-erío* or in *-ío*. This is the case for *guadería* ‘profession of a gaucho’ and *regrerío* ‘collection of ??’.

5.3.3 Italian: Collective nonce-formations in *itTenTen16*

Having considered French and Spanish, I will now turn to Italian, because as already mentioned, this Romance language seems to show interesting derivational patterns relevant to my research question regarding the derivational morphology of collection nouns. As shown in Table 5.1 and Figure 5.7, Italian has many suffixes that productively derive nouns from various domains. These can be semantically divided into three groups. As in the case of French and Spanish, Italian too shows a subdivision into primarily eventive (*-tura*, *-mento*, *-aggio*, *-ata*) and primarily collective nouns (*-ame*, *-ume*, *-aglia*). The third group in Italian is represented not by unproductive suffixes – every suffix studied here in fact derives a considerable number of nonce-formations – but by the suffix *-eria*, which does not show any semantic preference but equally derives nouns denoting events, collections and locations.

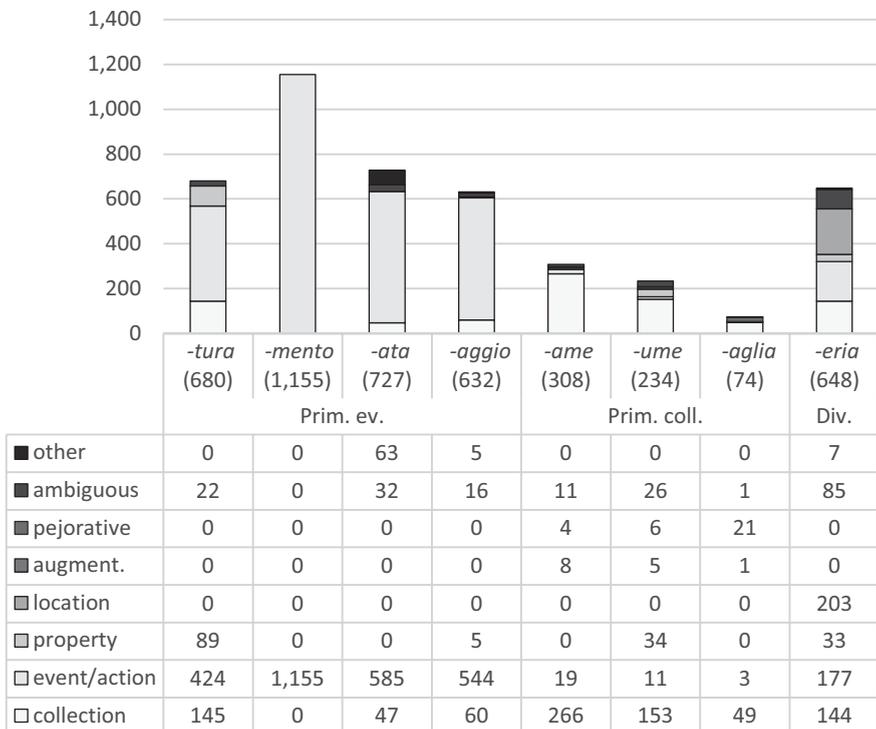


Figure 5.7: Semantics of unique nonce-formations in *itTenTen16* (1,000-word sample extrapolated to all hapaxes).

As shown in Figure 5.7, the Italian suffixes in general – except *-eria* – show a lesser degree of polyfunctionality than the suffixes investigated in French and Spanish. This is very apparent for the predominantly process-orientated suffixes *-tura*, *-mento*, *-ata* and *-aggio*. Nevertheless, these suffixes show clear differences with respect to their derivational base and their frequency in the corpus. As in the case of Sp. *-dura*, It. *-tura* e.g. also derives event nouns mainly based on verbs (cf. (63a); *svaligiatura* ‘robbing’ < *svagliare* ‘to rob’), whereas *-mento* and *-aggio* have various derivational bases (cf. (63b); *wipeoutaggio* ‘playing of the video game *Wipe Out*’) and *-ata*, like Sp. *-ada* is specialised for human proper nouns (cf. (63c); *bergoglionata* ‘typical action of Pope Francis’ < J. M. Bergoglio, civil name of the Pope).

- (63) a. E vanno anche interdetti tutti i politici e pubblici funzionari che hanno consentito la “*svaligiatura*”, ecc. (*itTenTen16*)
 ‘And all the politicians and the public officials, who agreed on the *robbing*, are going to be banned.’
- b. Siamo tutti abbastanza imbarazzanti rispetto ai tempi d’oro del *wipeoutaggio* giornaliero. (*itTenTen16*)
 ‘We are all quite embarrassing with respect to the daily golden days of *playing Wipe Out*.’
- c. L’aspettavamo. . . la *bergoglionata* quotidiana! (*itTenTen16*)
 ‘We waited for him . . . the daily *acting of Pope Francis*!’

The possible but marginal secondary function of these suffixes covers collection nouns in particular, although many of these are ambiguous. There are nevertheless clear examples, like in (64), where only an interpretation as a collection is possible. Those collection nouns derived upon primarily eventive suffixes are functional, i.e. denominal, and the derivational base is thus mostly represented by human or discrete inanimate nouns. (64a) e.g. is about fly fishing and the crafting of suitable baits for this sport. The two ad hoc constructs *palmeratura* (< *palma* ‘palm’⁶¹) and *anellatura* (< *anello* ‘ring’) designate the collection of feathers and small metal rings that constitute such a bait for fly fishing. Example (64b) illustrates a noun denoting a collection of human referents where *mandrappata* ‘bunch of prostitutes’ is derived from *mandrappa* (reg.) ‘prostitute’, in this context a synonym for politicians.

⁶¹ The exact derivational base of this ad hoc construct is unclear. From photographs and the context on the website, it becomes indeed clear that the description of the bait is about feathers used to make it. But ‘feather’ in Italian is not *palma*, but *penna*. An explanation could be that the feathers in this fishing bait are organised like the fan-shaped leaves of a palm tree.

- (64) a. Le mie cosiddette elk hair caddis sono fatte con corpo esile e color oliva, la *palmeratura* con gallo rosso ha barbule corte e l'*anelletura* ben distanziata, [. . .] (*itTenTen16*)
 'My so-called elk hair caddis are made of a delicate body and olive colour, the *whole of the feathers* is of red cock with short barbules and the *rings* well separated, [. . .]'
- b. non saremmo forse noi un triste paese governato da una *mandrappata* di vecchi decrepiti! (*itTenTen16*)
 'wouldn't we perhaps be a sad country governed by a *bunch of prostitutes* of old dodderers!'

Before coming to the predominantly collectivising suffixes in Italian, the highly polyfunctional *-eria* will firstly be examined. As shown by the analyses and overviews made so far, this suffix covers many semantic domains in every Romance language being studied in this present framework. This is not only the case for already lexicalised nouns, but also for ad hoc constructs. In Italian, *-eria* equally covers the semantic domains of collections, events and locations. Beginning with this last domain, *-eria* mostly derives names for restaurants, the base noun representing the food served there. Examples of this kind are *spuncetteria* (< *spuncetti* 'Venetian finger food'), *michetteria* (< *michette* 'small panini') or *ripeccateria* (< *peccato* 'sin'), this last ad hoc construct perhaps has to be understood as 'restaurant, where sinful, i.e. fat, dishes are served'.⁶² The eventive nonce-formations mostly represent the nominalisation of an action connected to a certain noun or adjective. In (65a) e.g. the property of being a *piagnone* 'cry baby' is converted into (permanently) acting like one. The collection nouns based on *-eria* are again mostly functional and not spontaneous and thus correspond to collections of mostly inanimate referents. The example of *golfetteria* in (65b) consequently designates a plurality of *golfetti* 'polo shirts'. In contrast, examples like (65c), where the ad hoc construct is derived from a proper name, are less frequent in my sample. In this case, the fashion brand *Gucci*, here with the (supposedly pejorative) alternative spelling *cuggi*, is taken as a centre of attraction and transfers

⁶² The exact derivational path is not clear in this case. It may either be based on the past participle of *ripeccare* 'to sin again' (*ripeccato* > *ripeccateria*), or it may be based on the noun *peccato* 'sin'. In this case, the prefixation with *ri-* is assumed to happen only after the suffixation with *-eria* (*peccato* > *peccateria* > *ripeccateria*). This latter option then has to be taken as a pun. In this sense, it is similar to other puns related to locatives like the coffeehouse franchise *Caffè Ritazza*, where the noun *tazza* 'cup' is also prefixed with the verbal iterative *ri-*, in the sense of cups that can be refilled (I am thankful to S. Dessi Schmid for this helpful comment).

all the properties associated with it and people wearing it to a contextually determined collection of people, *la cuggineria*.

- (65) a. Natalie Portman è toccante, ma senza strabordare nelle espressioni di facile *piangioneria*. (*itTenTen16*)
 ‘Natalie Portman is touching, but without overflowing in the expressions of *easy howling/acting like a cry baby*’
- b. Il commercio al minuto di articoli di abbigliamento, *golfetteria*, calze, camicie, cravatte, telerie confezionate, sottovesti e coperte. (*itTenTen16*)
 ‘The retail trade of clothing, *polo shirts*, socks, shirts, ties, fabric, petticoats and blankets.’
- c. [. . .] “tea-cup” dogs la moda dei cani miniaturizzati di valeria rossi “ho uno yorkshire toy” è sempre stata una della frasi-simbolo della “*cuggineria*” popolare in tema cinofilo. (*itTenTen16*)
 ‘[. . .] “tea-cup dogs”, the fashion for miniature dogs of Valeria Rossi “I have a Yorkshire toy” has always been one of the symbolic phrases of the popular “*cuggineria*” (*i.e. people typically wearing Gucci*) on the subject of dogs.’

The two examples of *golfetteria* and *cuggineria* could be cases of ad hoc constructs that are OMNs. Both nonce-formations designate an unspecified mass of polo shirts and human beings respectively and thus correspond to the typical properties of an OMN. My impression is that *golfetteria* and *cuggineria* would thus not be compatible with e.g. the indefinite article. These examples contrast with *mandrappata*, *palmeratura* and *anellatura* in (64), where the collective ad hoc constructs instead represent bounded collections. I will address the question of (un)boundedness of collective nonce-formations in chap. 5.3.4 in more detail.

Turning finally to the group of primarily collectivising suffixes in Italian, *-ame*, *-ume* and *-aglia*, a first observation is that their quantitative relation to the primarily eventive suffixes, *-tura*, *-mento*, *-aggio* and *-ata*, is similar to that found in the Spanish data. It can thus be stated that these suffixes are less productive than those that mainly derive action nouns but that the frequencies are still considerable. The three Italian collectivising suffixes differ, however, from their Spanish equivalent in deriving nearly exclusively collection nouns. In this respect, *-ame* mostly derives denominal collection nouns designating pluralities of discrete inanimate or human referents. As illustrative examples, *gioellame* in (66a) designates a plurality of *gioielli* ‘jewels’ and *vecchiame* in (66b) refers to a collection of *vecchi* ‘elderly people’. Particularly this last example shows that these ad hoc constructs on *-ame* often have an additional negative connotation. This does not necessarily have to do exclusively with the suffix, as explained in

chap. 3.2, but rather is due to an interplay between the suffix and the semantics of the base noun. The human referents in (66b) e.g. have no pejorative connotation, but grouped together with the aid of the suffix, the individual character of each person is backgrounded and thus devalued. This becomes even clearer when the ad hoc construct *vecchiamè* is contrasted with the discrete plural *vecchi*. Other examples in the corpus already have a negatively connoted base noun, often devalued through context. Nonce-formations like *riccame* ‘bunch of rich people’ (< *ricco* ‘rich’) in (66c) or *giudeame* ‘bunch of Jews’ (< *giudeo* ‘Jew’) in theory have no pejorative base, but in that particular discourse, rich and Jewish people are already seen negatively, an effect which is then reinforced through the suffixation.

- (66) a. In quella barca, infatti, c'erano investiti un mucchio di soldi, ma in opere d'arte non in futile *gioellame*. (*itTenTen16*)
 ‘In that boat, in fact, there was invested a lot of money, but in works of art not in a futile *collection of (cheap/false) jewels.*’
- b. [. . .] mentre gli altri suoi organi di alto livello erano stati giusto una discarica da riempire con *vecchiamè* inutile, il cui potere era solo di chiacchierare. (*itTenTen16*)
 ‘[. . .] while his other high-level organs had been just a dump to fill with useless *old people*, whose only power was to chat.’
- c. Violentando la natura, distruggendo l'ecosistema, gentrificando i quartieri, spogliandoli della presenza inutile di operai, artigiani, stranieri, studenti, artisti, a favore del *riccame ignorante e scintillante*. (*itTenTen16*)
 ‘Violating nature, destroying the ecosystem, gentrifying neighborhoods, stripping them of the useless presence of workers, artisans, foreigners, students, artists, in favor of the *bunch of ignorant glitterati.*’

In contrast to these only potentially pejorative ad hoc constructs, which of necessity are collective but are only negatively connoted as an option, the suffix *-aglia* also has the function of devaluating the base without a collective meaning. Like French *-aille*, It. *-aglia* adds a negative connotation to a base like *Banca d'Italia* in (67a). Here there is no sense of plurality but only of devaluation due to the suffix. The pun on *Italia* is indeed well represented in my sample. I found 11 occurrences of this type (cf. e.g. *Alitaglia* < *Alitalia*, an Italian airline, *dis-equi-taglia* ‘(Italian) imbalance/inequality’ < *disequi-* ‘unequal-’). Nevertheless, I also found collective ad hoc constructs on *-aglia* in my sample. These can, like the nonce-formations on Fr. *-aille*, have a negative connotation, based on pejorative as well as on neutral bases, as well as no connotation at all. In (67b) e.g. there is another case of human referents who, without any context, do not have to have any connotation at all (*profugo* ‘refugee’), but given a certain context and the pluralisation

with *-aglia*, they become devalued (*profugaglia* ‘bunch of refugees’). Examples like (67c), where a pejorative interpretation is not possible, can however also be found in the corpus. Here the base *gazza* ‘magpie’ is simply pluralised in the sense of ‘population of magpies’. Whereas It. *-ame* thus derives mostly neutral ad hoc constructs, It. *-aglia* is specialised for pejorative nonce-formations, which can also refer to more than one referent, but do not have to.

- (67) a. L'oro della *bancaditaglia* pagato con il lavoro e le tasse degli italiani non sta piu in italia da tanto tempo [. . .] (*itTenTen16*)
 ‘The gold of the damn “Banca d’Italia” paid with the work and taxes of Italians is no longer in Italy for a long time [. . .]’
- b. si sciorinano dati sulla percezione dell’insicurezza di noialtri italiani, io e lei, a causa dei merdosi invasori mori, della sedicente *profugaglia* . . . a quanto stiamo? (*itTenTen16*)
 ‘we are collecting data on the perception of insecurity of us other Italians, you and I, because of the damn invading Moors, the self-styled *bunch of refugees* . . . How much are we up to?’
- c. Dopo il ponte di Ca’ Venier, si arriva al Lago di Ferro, che ospita la più grande *gazzaglia* del Delta, in cui nidificano garzette, aironi bianchi e rossi, falchi di palude. (*itTenTen16*)
 ‘After the bridge of Ca’ Venier, we arrive at the Lago di Ferro, which is home to the largest *population of magpies* of the Delta, where egrets, white and red herons, marsh harriers, nest.’

The analysis of ad hoc constructs in *itTenTen16* cannot confirm the assessment of Magni (2018, 217), who argues that the suffix *-aglia* is especially prone to proper nouns. In my sample, I found 36 collective ad hoc constructs, of which only a single nonce-formation is based on a proper noun. This is the case of *raffiche di mitrokinaglia* ‘gusts of everything having to do with the person Mitrokhin and his archive’. The nonce-formation is based on the proper name of *Vassili Nikitich Mitrokhin*, who made numerous handwritten notes of the Soviet Union during his colonelship in the KGB and who defected to the west with this archive.

The suffix *-ume* finally shows a behaviour similar to *-ame*. It particularly derives collection nouns of discrete inanimate or human referents, which may or may not have an additional pejorative connotation. The latter case is illustrated in (68a), where *oggettume* ‘collection of objects’ (< *oggetto* ‘object’) simply designates the various objects in a tabletop game to mark positions of the game-defining elements, without any additional evaluation. When the ad hoc constructs are pejorative, the devaluation is mostly not just effected through the suffixation but by means of the whole context. This can be seen in (68b), where *benpensante*

‘trusting/benign’ is not necessarily negatively connoted, but it is negative in this particular context – the suffixation with *-ume* again reinforces this (*benpensantume* ‘bunch of do-gooders’). As in this case of *benpensante*, it is often unclear whether the base of the ad hoc construct on *-ume* is an adjective or a noun, both options are possible (cf. also examples like *democraticume* ‘bunch of democrats’ < *democratico* ‘democratic’). This confirms what Rainer (2018b, 420) already states for lexicalised derivatives on *-ume* like *frittume*, based on *fritto* ‘fried/fried food’. This etymologically established orientation towards states and properties (cf. Table 5.1) is thus still present in more recent nonce-formations on *-ume*. In the sample of *itTenTen16*, there are various ad hoc constructs like the one in (68c), which only license an interpretation as a state or property, never as a collection.

- (68) a. Al centro del tavolo vi e’ uno spazio ove vengono poste le miniature o altro *oggettume* vario che mostreranno la posizione dei personaggi in relazione agli altri, agli elementi del paesaggio e a eventuali nemici. (*itTenTen16*)
 ‘In the middle of the table there is a space where miniatures are placed or other various *objects* that will show the position of the characters in relation to others, other elements of the landscape and any enemies.’
- b. Una storiaccia di alcolizzati, morti, stupri, *benpensantume* vario. (*itTenTen16*)
 ‘A bad story of alcoholics, deaths, rapes, various (*damn*) *trusting people*.’
- c. Ancora 12 giorni e potrò misurare la curiosità mi sta mangiando viva per il resto i capelli continuano a giovare in lucentezza e *lisciume* e setosità. (*itTenTen16*)
 ‘Still 12 days and I will be able to measure, the curiosity is eating me alive, for the rest, the hair continues to benefit in lustre and *smoothness* and silkiness.’

This detailed analysis of the Italian data can be summarised as follows: the unique nonce-formations in *itTenTen16* based on the suffixes of Table 5.1 can be subdivided into two groups: those that are clearly process-orientated and those that are clearly collective. The semantic range of nonce-formations in these groups is generally limited to these two main functions. Although there are secondary functions, these are quantitatively marginal – especially compared to their proportions in French and Spanish. The suffix *-eria* does not show any semantic preference and equally derives event nouns, collection nouns and locatives. The two semantic domains of processes and collections thus make up two clearly delimited derivational groups in Italian. Collective suffixes in Italian are less productive than eventive suffixes but they still cannot be considered as unproductive.

5.3.4 Pan-Romance comparison

After the detailed analysis of the ad hoc constructs in the three languages examined here, I will in this last section address pan-Romance similarities and differences. Recapitulating what has been elaborated so far, the underlying question of this chapter on morphology was to analyse the extent to which the aspectual category of collection nouns is constituted not only by semantic-syntactic aspects but also by derivational criteria. Basing the treatment of this issue generally on a constructionist approach and having summarised the state of the art on collective suffixes, this chapter addressed the central question by analysing unique nonce-formations. As has been mentioned various times, there are indeed many similarities between the three Romance languages studied here, but also differences and incongruencies.

Beginning with the similarities, several suffixes display analogous behaviour in two or even all three of the Romance languages examined. This is the case of Fr. *-erie*, Sp. *-ería* and It. *-eria*, which in all three languages have a more or less even semantic distribution over events, collections and locations. The continuations of Lat. *-ATICUS* and *-ATA* also have a similar functional distribution in all three languages in being (nearly) exclusively restricted to deriving event nouns. In addition, every language studied here seems to have one more or less productive suffix that predominantly derives collection nouns. This is the case with Fr. *-aille*, Sp. *-erío* and It. *-ame*, *-ume* and *-aglia*. These suffixes derive mostly functional and not spontaneous collection nouns, i.e. the collective ad hoc constructs based on them are mainly denominal and they thus directly group together a plurality of underlying referents (cf. Baldinger 1950, 215). In this respect, the numerous nonce-formations based on proper nouns have to be mentioned as an exception. These are de facto not functional collection nouns, because they do not simply group the referents of the base noun together. Instead, they show linguistic characteristics similar to associative plurals in taking the proper noun as a centre of attraction and in grouping together a heterogeneous quantity of referents associated with this proper noun (cf. Mauri 2017, 310–311).

Furthermore, various collective ad hoc constructs have the same derivational base but are derived with different suffixes. This is the case with Sp. *obisperío/obispamen* ‘collection of bishops’, *mierdaje/merderío* ‘collection of shit’ and *banderaje/banderamen* ‘collection of flags’ as well as It. *pellatura/pellaggio* ‘collection of drumheads’ and *servame/servume* ‘collection of servants’. This goes in line with the assumption of Magni (2018, 214), who states that the collective derivational suffixes *-ame*, *-ume* and *-aglia* in Italian are to some extent synonymous, as in *ferrame/ferraglia* ‘scrap metal’. I did not find cases like these in the French sample, the example of *cambrousserie/cambroussaille* mentioned above do not share the same meaning.

I have already mentioned the relation between COLLECTIVITY and pejoration. In this respect, it has been generally assumed that every collective suffix has at

least the potential ability to connote the whole meaning of a derivate negatively (cf. Baldinger 1950, 218–222; Rainer 1993, 207; Grossmann 2004, 245). As in the case of appreciative derivation, collectivising also either reduces the perceived importance of an individual by using diminutives, for example, or amplifies in the sense of ‘more of the same’, with the aid of augmentatives. All three processes “may add vagueness and also devaluation to the meaning of the derivative” (Magni 2018, 221; cf. also Merlini Barbaresi 2004, 292). This connection between COLLECTIVITY and pejoration may be so strong in certain suffixes that they have the ability to derive ad hoc constructs that have an exclusively pejorative connotation. This is especially the case with It. *-ume* and *-aglia* as well as Fr. *-aille*. As described in the respective chapters, many of the negatively connoted neologisms have a neutral derivational base, and the pejoration thus mainly comes from the use of the suffix and possibly arises also from the statement’s context. In contrast, there is also a wide variety of collective nonce-formations that are not negatively connoted. This is not only the case with those numerous collection nouns that group together inanimate objects, but also with those that combine human referents. In (69), some examples from Italian, French and Spanish illustrate this.

- (69) a. con el despotismo político apoyado en todas partes en la religión como dogma, en la iglesia como autoridad, en el clero y *frailerío* como fuerza, [. . .] (*esTenTen11*)
 ‘with political despotism supported everywhere by religion as dogma, by the church as authority, by clergy and *all the monks* as force, [. . .]’
- b. “E visto che è anche una mamma,” continuò Giulia, “l’orchessa nasconde Pollicino e i suoi fratelli sotto il letto per proteggerli da quell’orco del marito, un bonaccione, in realtà, ma goloso di *bimbame*. (*itTenTen16*)
 ““And since she is also a mother,” continued Giulia, “the wife of the monster hides the Little Thumbling and his brothers under the bed to protect them from her ogre husband, a good-natured man, in fact, but a devourer of *children*.”
- c. Il bricole des armes, les répare, entretient la vieille fraise de dentiste de la *guérisserie*, redresse le zinc du vieil arrosoir. . . (*frTenTen12*)
 ‘He crafts weapons, repairs them, maintains the old dentist’s drill of the *healers*, straightens the zinc of the old watering can. . .’

The comparison of the data analysed thus confirms the obvious relation between COLLECTIVITY and pejoration, which is stronger in some suffixes than in others. It adds a piece to the puzzle that the pejorative feature of many collection nouns is not necessarily a result of semantic change but already exists in the functional range of the respective suffixes themselves. The connection thus can be present,

but does not necessarily have to be activated and is therefore to be understood as a semantic potential, possibly activated by pragmatic aspects.

Before coming to the treatment of the aspectual types of collection noun present in the data of the *TenTen* corpora analysis, a short note on the referents in general: in the present analysis, human as well as discrete inanimate referents were mentioned as the most common referent classes in our sample. I also found animate referents among the bases of the ad hoc constructs analysed (e.g. Sp. *gavioterío* ‘swarm of seagulls’ < *gaviota* ‘seagull’, cf. also (67c)), but these are rather rare and the nouns used often do not refer to the animals they denote but to human beings, and therefore have a devaluating connotation (e.g. It. *badri-pame* ‘bunch of lazy people’ < *badriipo* ‘sloth’). There are also some instances of collective ad hoc constructs based not on count nouns but on mass nouns. In these cases, there is not just only a bounding process triggered by the suffixation, but two more steps are necessary to understand the meaning of the ad hoc construct: For instance, in the case of Sp. *ziquerío*, there is firstly a metonymic shift of the meaning of the base noun of Sp. *zinc* ‘zinc’ to ‘object made of zinc’. Only then may the meaning of the suffix *-erío* be compatible with the meaning of the base noun and the individual object in the denotation may first be pluralised and then bounded. The derivational result of *ziquerío* then denotes a coherent collection of objects made of zinc. Another example is shown in (70), where the former mass noun is first bounded and then pluralised and then this plurality is again bounded. In this respect, Sp. *moquerío* (< *moco* ‘slime’) refers to the bounded quantity of various portions of slime which are in the lung:

- (70) Si no se saca la flema y el *moquerío* los pulmones se van a infectar.
(*esTenTen11*)
‘If one doesn’t get the phlegm and the *portions of slime* out, the lungs will get infected.’

In the analyses of the individual languages, the question of whether the collective ad hoc constructs are CCNs or OMNs could not be addressed, and I only applied the minimal rule of internal plurality to categorise an ad hoc derivate as a collection noun or not. To now address this issue, I examined whether the ad hoc construct under study occurs in a syntactic count or mass context, i.e. whether it is presented in the discourse as bounded or unbounded. The only systematic criterion for categorising these contexts in Romance is determination, so I classified the determiners of the collective ad hoc constructs as either count, mass or ambiguous (cf. Table 1.2). A more detailed analysis of semantic aspects such as the accessibility of the constituting individuals is not possible in this framework, due to the often very limited range of context available. As already indicated in connec-

tion with the overall description of syntactic marking of aspectuality in Romance languages in chap. 1.2.2, the indefinite article as well as numerals are clear indicators for aspectual boundedness. In addition, distributive quantifiers like *every* (Fr. *tout/chaque*, Sp. *todo/cada*, It. *tutto/ogni*) or *various* (Fr. *plusieurs, quelque(s)*, Sp. *vario(s), alguno(s)*, It. *vari(o), qualche, parecchi(o)*) are considered as being able to bound the noun in question aspectually (cf. Kleiber 1998; 2012, 220–221; Flaux 1999, 480). However, a clear indicator for mass syntax can only be found in French and Italian, where it is represented by the partitive article. Finally, quantifiers like *more, less, (too) much, enough of, all* as well as the definite article, possessive pronouns, demonstratives etc. are not specified for the mass-count distinction and are therefore ambiguous. The results of this analysis are displayed in Figure 5.8, where I included in this analysis not only collective ad hoc constructs but also ambiguous cases; the numbers in the rows indicate the absolute frequencies of occurrences.

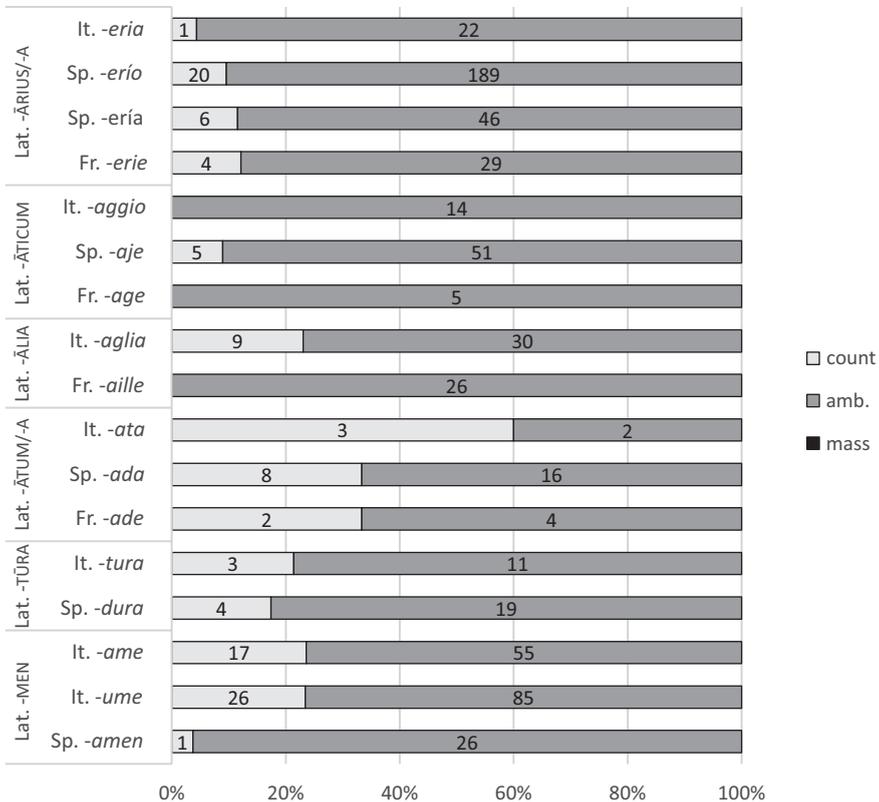


Figure 5.8: Distribution of mass and count determiners in combination with collective ad hoc constructs.

As illustrated in Figure 5.8, the great majority of determiners occurring in combination with collective ad hoc constructs in French, Spanish and Italian are ambiguous with respect to the mass-count distinction. One can further detect a comparatively high proportion of clear count contexts for the continuations of Lat. -ĀTUM/-A, but apart from that, a mere analysis of the determiners cannot address whether a particular suffix predominantly derives CCNs or uncountable OMNs. There are no instances of a collective ad hoc construct that is determined by the partitive article, and for this reason there are no cases of mass syntax indicated in Figure 5.8. Nevertheless, this complete absence of mass contexts does not necessarily lead to the assumption of a total lack of mass nouns in my sample. Even in the case of prototypical mass suffixes like Fr. *-aille* (cf. *supra*), I did not find any occurrences of the partitive article.

Although there are indeed unambiguous instances of count contexts, the collective nonce-formations in the *TenTen* analysis generally seem to lie in between the two poles of bounded groups and mass categories. Examples of the former kind are Fr. *umpèterie* (cf. (56b)), Sp. *banderaje* (cf. (60a)), Sp. *tumberío* (cf. (61b)), It. *palmeratura/anellatura* (cf. (64a)), and It. *oggettume* (cf. (68a)). In examples like (71a), the suffixation bounds the referents of a lexicalised OMN which serves as a derivational base. In this sense, *genterío* is a bounded group of *gente* ‘people’, i.e. the crowd. In contrast, many of the collective ad hoc constructs in the data set cannot be clearly categorised as either mass or count. Classifier constructions, which may indicate usage as an OMN (cf. chap. 1.2.2), like It. *articoli di golfetteria* in (65b), are scarce. Neither are there clear cases where e.g. a hyponymic relation would exclude the possibility of a CCN. Constructions like *il resto dell’* in (71b) possibly imply a typical CCN constituted through meronymic relations. This is however only an optional categorising criterion since OMNs may also be part of a meronymic collection-member hierarchy (cf. chap. 3.1.2 and chap. 4.2.2).

- (71) a. Dave Gahan aparece bailarín, el *genterío* grita por él (hombres y mujeres). (*enTenTen11*)
 ‘Dave Gahan appears dancing, the *crowd* screams for him (men and women).’
 b. Tutto il resto dell’*alfabetume* sportivo inneggiava al gioco “aperto”. (*itTenTen16*)
 ‘The rest of the *sports journalists* (*lit. literates*) praised the “open” game.’

Nevertheless, taking into consideration the pragmatic mechanisms underlying such derivation of collective ad hoc constructs, the impossibility of clearly categorising the constructs becomes understandable. The majority of contexts in the

TenTen corpora are not necessarily negative as described above, but often expressive, and this is shown by the quotation marks often found in the examples, typical for creative nonce-formations (cf. (58c); (60d); (63a); (65c); cf. Dal/Namer 2018, 206–208). In using suffixation as a means of expression, the speaker not only pluralises the referents but always adds something, be it simply an external boundary or some connotation. This is most clear in the numerous nonce-formations based on proper names, mostly of politicians (cf. also Huguin 2018). These collections are held together particularly through contiguity, they are always interpreted in a specific semantic frame and thus are very similar to CCNs. Here the exact members also depend on the specific context, and they do not share inherent semantic features. In this respect, the hearer/reader of collective ad hoc constructs like It. *berlusconame* (cf. Mauri 2017; Magni 2018), Sp. *Fujimontesinada* (cf. (60b)) or also Fr. *sarcosaille* ‘adherents of Nicolas Sarkozy’ (*frTenTen12*) has to be familiar with the specific frame and its constituting features to be able to interpret them correctly. Without knowing of the escapades of Silvio Berlusconi or the corruption affairs of the former Peruvian president Fujimori and his secret service chief Montesinos, one could not ascribe any characteristic to the groups of people designated by the derivatives. On the other hand, however, the exact extension of the ad hoc construct is blurred by the massifying of the referents. “La sarcosaille en délire” (‘the adherents of Nicolas Sarkozy in a state of delirium’) does not denote a clearly delimited group of persons, but rather a de-individualised mass (cf. also Mauri 2017, 319). As a consequence, many collective ad hoc constructs are in fact not prototypical CCNs, but neither are they typical OMNs – they form an intermediate category. Part III will further address the hypothesis that OMNs may only come into being by diachronic evolution. This assumption is supported by research on Engl. deverbal conversion and nominalisations based on *-ing*. For referential nominalisations, Andreou/Lieber (2020) did not find any correlation between the type of derivational process involved and the syntactic mass or count properties of the derived noun. More importantly, the context or certain modifiers seems to play a role for the actual interpretation of the derivative. There seem to be tendencies of particular derivational processes to either favour mass or count noun derivations, but similar to this present analysis, they also found doublets like *finishings*_{count} and *finishing*_{mass}. The results of Andreou/Lieber (2020) as well as my own analysis indicate that it is not the suffix which determines the status of the derivative as a mass or count noun, but rather pragmatics and – as we will see in part III – diachronic evolution. It would be fruitful to investigate further the actual transparency of collective ad hoc constructs, as well as their syntactic and semantic properties, by testing them with native speakers of French, Spanish or Italian. Unfortunately, this further analysis goes beyond the scope of the present thesis.

Completing this section on the derivational processes underlying collection nouns, I can now come back to the initial research question, repeated here for convenience:

RQ_{morph}: To what extent does the realised productivity of collective derivational processes reflect their expanding productivity shown in the number of newly coined words based on them? To what degree can the category of collection nouns then be defined as a sharply delimited derivational category?

Having considered not only the already lexicalised collective (micro-)constructions (realised productivity), but also newly derived collective ad hoc constructs (expanding productivity), I can affirm the initial supposition that collection nouns do indeed represent a derivational category in Romance. The assumed constructional schema <collection> can be formalised as follows, where Y represents an abstract suffix-variable and X a variable of the underlying derivational base noun:

$$\langle [[X]_N -Y] \leftrightarrow ['\text{plurality of } X']_N \rangle$$

This schema is at a very high level of abstraction which serves, at this point of my elaboration several goals. First, it leaves open the possibility of systematic integration in a network of polyfunctionality and thus the connection to other secondary functions taken on by the particular suffix. Second, it also leaves open the particular aspectual type of collection noun represented by a construction or construct, whether this is a CCN or an OMN (cf. Figure 3.1). For these reasons, the various aspects discussed in the language-specific sub-chapters are summarised in what follows to then enable a semantic map of COLLECTIVITY in the Romance languages to be drawn.

As already indicated in the description of the individual languages, it seems to be the case that the morphological category of collection nouns in French is strongly dependent on the derivational category of processes, whereas in Italian both derivational categories are independent from one another. Spanish seems to lie somewhere between these two poles, with *-erío* being more or less independently collectivising, but the rest of the suffixes being more attached to the derivational category of event nouns. This dependence or independence mostly manifests itself in the degree of polysemy of the particular suffixes. In French, the suffixes in general are highly polyfunctional whereas in Italian it could be stated for every suffix that secondary functions of the particular suffixes are quite marginal. As already outlined in chap. 5.3.1, it seems to be the case that because of the phonetic erosion of suffixes in French, the overall number of possible suffixes has decreased. As a consequence, either the still existing suffixes have to

take over more functions or speakers have the possibility of choosing an analytic means of expressing concepts like DIMINUTION or COLLECTIVITY. This is not the case in Spanish or Italian (cf. Table 3.1). Nevertheless, my empirical analysis revealed that each of the three Romance languages examined has at least one suffix that predominantly derives collection nouns. Differences between French, Spanish and Italian, however, also became apparent, with French being the language with the least productive derivational pattern (only one suffix, comparatively low quantity of ad hoc constructs) and Italian with the most productive derivational pattern (three suffixes, comparatively high quantity of ad hoc constructs). Spanish was said to function as an intermediate case (only one suffix, comparatively highly productive). There is thus indeed a constructional schema <collection> as illustrated above, upon which speakers of the different Romance languages can interpret already existing derivatives, but upon which they are also able to derive new forms. This has been evidenced by the empirical analysis of ad hoc derivatives. I am now able to sketch a semantic map of the functional range of collectivity in Romance which illustrates the semantic relations and connections as well as the diachronic paths of evolution elaborated in chap. 3.2. This map is represented in the following Figure 5.9:

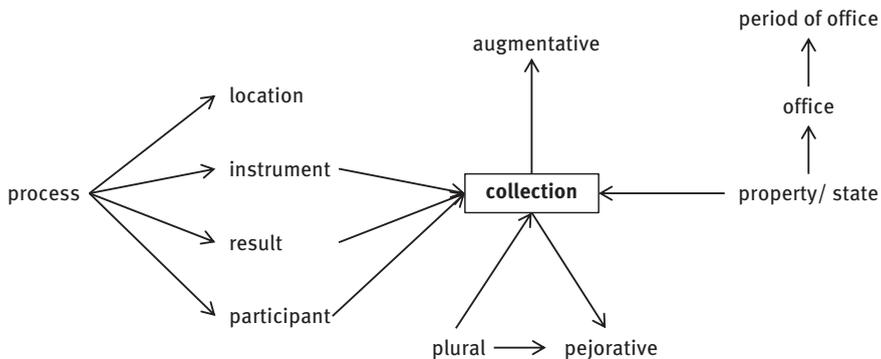


Figure 5.9: Semantic map of collectivity in Romance.

As illustrated in Figure 5.9, there are three possible diachronic collective pathways, which are assumed to be not only particular to Romance languages, but perhaps also to be more or less universal. First of all, there is the path starting from processes and deriving collection nouns through the results, instruments or participants involved in these processes. This is the case with the Latin formerly eventive suffixes *-MEN*, *-TŪRA* and *-MENTUM*. Second, there is the path which starts from denoting a property and then transferring this property to the col-

lection. This is the case of the former Latin suffixes *-ĀRIUS* and *-ĀTICUS*, which derived relational adjectives. And finally, there is the pathway that leads from the Latin neuter plural to a collection noun. In addition to these pathways leading to the centre represented by the function of expressing COLLECTIVITY, one could theoretically expand this map even more to other neighbouring functions. This is exemplified by the expansion of PROPERTY/STATE to OFFICE and PERIOD OF OFFICE, e.g. represented by the Sp. suffix *-ario* (cf. Table 3.1). Starting from the function COLLECTION, there are two additional functions, first the already fully elaborated connection to pejoration and second, the relation to augmentatives, e.g. represented by Sp. *-amen*. This map covers all possible semantic functions connected to COLLECTIVITY and thus is able to represent this derivational category accordingly. The semantic range of this derivational category is finally represented in Figure 5.10.⁶³

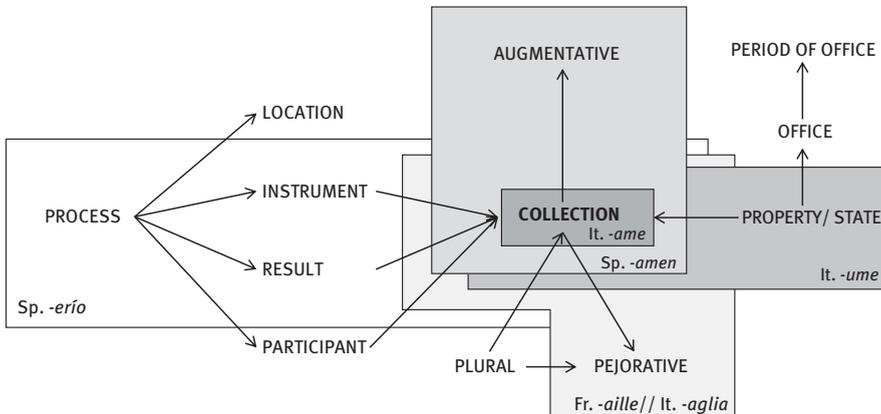


Figure 5.10: Semantic map of the functional range of collective suffixes in Romance.

Every suffix represented in this map has a predominantly collectivising function as has been analysed both by means of the lexicographic information available and by the nonce-formations. In this respect Fr. *-aille* and It. *-aglia* additionally derive pejoratives, It. *-ume* properties and states, Sp. *-amen* augmentatives and Sp. *-erio* processes. It. *-ame* in fact has the exclusive function of collectivising underlying referents. In conclusion, it can be shown that the derivational processes upon which collection nouns are derived in the Romance languages

⁶³ The various shades of grey should help to distinguish the different domains of the suffixes and have no further categorical meaning.

studied here constitute another definitional criterion of this aspectual category. In some languages, the collective function is more exclusive than in others, but it is always present. What this semantic map cannot show are further pieces of information about the different suffixes, like the different types of collection noun derived, i.e. count vs. mass, prototypicality effects and so on. As thematised in chap. 5.1, the representation of these aspects is not the intention of a semantic map as elaborated by Haspelmath (2003). These may and must be specified in the description of both the constructional schema and the corresponding semantic map. A possible representational method could be, for instance, the reflection of quantitative tendencies by using spider diagrams. This present analysis not only adds another piece to nominal aspectuality puzzle, it could also help to refine lexicographic information on the category of collection nouns and the derivational processes involved.

6 Conclusion: Linguistic characterisation of Romance collection nouns in present-day language

Part II of this present thesis analysed the linguistic characteristics of French collection nouns in comparison to Italian, Spanish and Portuguese in the synchrony of present-day language. The analyses aimed to examine research questions 1 and 2 formulated in the introduction:

RQ1: What are the influencing factors on the particular linguistic expression of a collection of entities and the semantic-syntactic characteristics related to it?

RQ2: To what extent are there any productive word-formation patterns in the domain of COLLECTIVITY?

The first research question (RQ1) refers to the semantic-syntactic expression of COLLECTIVITY in Romance languages. Given the theoretical framework of nominal aspectuality, the first influencing factor concerns the linguistic construal of extra-linguistic entities as different aspectual types correlating with different nominal categories, *viz.* count collective nouns as well as singular and plural object mass nouns (cf. RQ_{ss}1). It was assumed in this respect that the nominal type verbalising a collection reflects its conceptual construal. The state of the art summarised in chap. 3.1 pointed to a continuum of these three main types of collection noun which is governed by the accessibility of the internal plurality of a collection as well as by its external boundedness. The two factors are assumed to run in opposition to each other: an increasing accessibility of the plurality of constituting entities implies a decreasing of the external boundedness of a collection, and vice versa (cf. HP_{ss}1). This continuum was confirmed by the acceptability judgement tests conducted: CCNs, which are assumed to have a high degree of external boundedness and a low degree of accessibility of the internal plurality of the collection, are not at all compatible with such phenomena as stubbornly distributive predicates and constructions like ‘one after the other’. SOMNs overall show a lesser degree of compatibility with linguistic means of expressing countability and a higher degree of compatibility with distributive predicates. POMNs are even more compatible with distributive predicates, but still not as countable as object nouns. The second semantic-syntactic research question (RQ_{ss}2) then addressed the question of cross-linguistic differences in the compatibility of OMNs with linguistic means of expressing distributivity. According to the second semantic-syntactic hypothesis (HP_{ss}2), this compatibility correlates with

the degree of grammaticalisation of the mass-count distinction, summarised for Romance languages in Table 1.2. This correlation could be confirmed only to a certain degree. With respect to the countability of OMNs, the acceptability judgement tests confirmed a continuum of acceptability of OMNs in count-syntax going from French, to Italian, Spanish and Portuguese only for the semantic domain of CLOTHING (cf. tests 1a and 1b). Test 2a showed that OMNs are acceptable with stubbornly distributive predicates and are not acceptable with highly distributive predicates irrespective of the language and its particular nominal system – the assumed continuum could therefore not be confirmed for linguistic means of accessing the internal plurality of a collection. The degree of grammaticalisation of the mass-count distinction consequently only affects characteristics directly having to do with count syntax, whereas semantic aspects having to do with the internal plurality are not affected by it. The differences between the combinability of OMNs with stubbornly distributive predicates and highly distributive predicates could be explained by the quantifying operations done by these parallel to the predication operations: while stubbornly distributive predicates leave open the exact number of entities they predicate, constructions like ‘one after the other’ must quantify them. This operation is however not possible with OMNs since they leave open the question of the specific quantity of entities constituting them. Finally, the third semantic-syntactic research question (RQ_{ss3}) addressed possible influencing factors relating to the ontological type of entities making up the collection. It was assumed in this respect that an increasing animacy of referents goes hand in hand with an increasing saliency of them and consequently an increasing degree of accessibility in the collection (cf. HP_{ss3}). The acceptability judgement tests could not confirm any kind of animacy effect with respect to collection nouns and their semantic-syntactic characteristics. A little outside the scope of the first research question, the acceptability judgement tests also examined the question of whether the collections verbalised by OMNs are organised by a meronymic or by a hyponymic relation between the collection and its constituting entities. It could be confirmed that SOMNs are neither good hyperonyms nor good holonyms, but that they are nevertheless not impossible in constructions indicating such a status. For the POMNs tested, a relatively clear status as hyperonyms could be determined by means of the acceptability judgement tests.

Influencing factors on the particular linguistic expression of a collection of entities and the semantic-syntactic characteristics related to it are thus its construal by the speaker reflected by the nominal type he or she chooses, as well as the language-specific nominal system. In this respect, the expression of COLLECTIVITY is reflected by the specific lexical means of expression, i.e. the actual nouns, as well as the grammatical frame they are embedded in, e.g. the grammaticalisation of the mass-count distinction. With this, the empirical analysis of the

semantic-syntactic characteristics of collection nouns in Romance languages justifies the theoretical framework of nominal aspectuality as has been elaborated in part I, as well as the assumed parallelism to verbal aspectuality and in particular the onomasiological perspective adopted here and by Dessì Schmid (2014; 2019) amongst others for the verbal domain.

The second main research question (RQ2) addressed the derivational patterns collection nouns are associated with and in particular the question of whether collection nouns are a productive word-formation pattern in present-day French, Spanish and Italian. The analysis of collective nonce-formations in the *TenTen* corpora showed that in French the degree of expanding productivity of collective suffixes reflects their realised productivity: in this language, collection nouns mainly come into being by metonymic shift from e.g. event nouns. For Italian and to a lesser degree also for Spanish, it could be said, on the other hand, that a number of collective suffixes productively derive collection nouns, which refutes the assumption that the derivational pattern of COLLECTIVITY is synchronically not productive. Adopting the theoretical framework of construction morphology, the development of a constructional model of COLLECTIVITY was additionally possible, bringing together the assumptions made by construction morphology, e.g. default inheritance, and the intertwining of a network of collective suffixes and their possible meanings by recourse to the method of the semantic map.

III The diachronic development of French collection nouns

Having analysed the present-day characteristics of collection nouns in French in comparison to Spanish, Italian and Portuguese, the third part of the present work will focus on the diachronic evolution of this aspectual type. As already indicated, research done so far has already given quite a clear picture of the origins of collection nouns and plausible pathways of evolution. There are, however, no empirical investigations that have gone beyond the analysis of etymological dictionaries and quantitative corpus analyses. The last section of this monograph thus aims to empirically investigate the assumed pathways of evolution of collections. First of all, chap. 7 will give an overview of the state of the art on possible pathways of lexicalisation of collection nouns. These have already been indicated in chap. 3.2 and will now be addressed in detail. Findings from previous research studies will help me in formulating concrete hypotheses regarding the origins as well as the morpho-syntactic and semantic evolution of collection nouns in Romance. These hypotheses will be empirically investigated by means of an extensive quantitative and qualitative corpus analysis of French collection nouns (cf. chap. 8). Chap. 9 will conclude with a detailed description of French collection nouns in diachronic evolution.

7 State of the art: The assumed paths of lexicalisation of collection nouns

Most researchers assume some kind of pathway of evolution of collection nouns. This kind of linguistic change is often characterised by the notion of lexicalisation. In a first step, I will look at what kind of particular evolution is described by this concept and what kinds of path the development of nouns take. In a second step, I will concentrate on the state of the art regarding the lexicalisation of the specific aspectual type of collection noun. Here, I distinguish between the phylogeny and the ontogeny of collection nouns, examining the language-specific evolution of collection nouns in a speech community, as well as more or less language-independent findings on how collection nouns are acquired in a child's development. The former distinction can again be subdivided into traditional (cf. i.a. Baldinger 1950) and cognitive historical linguistic approaches (cf. Mihatsch 2006; 2016) as well as a particular line of enquiry arising from research on the mass-count distinction (cf. i.a. Grimm/Levin 2011). Finally, from this overview of the state of the art, I will deductively formulate research questions and hypotheses that will in turn lay the groundwork for the analysis of the corpus examination set out in chap. 8.

7.1 The concept of lexicalisation

The following elaboration of the concept of lexicalisation is organised in two sections. Since the notion of lexicalisation is traditionally associated with the idea of grammaticalisation, I will first touch upon the various approaches to lexicalisation in the history of linguistics. This section will therefore focus on the parallels between and differences relating to grammaticalisation and language change discussed in the literature. The second section will focus on the understanding of lexicalisation in recent research, focusing on definitional aspects independent of grammaticalisation. The traditional approaches to lexicalisation that concentrate mainly on the formal aspects of this process, e.g. univerbation, will be argued to be less important for the development of collection nouns. In contrast, semantic processes such as typical paths of meaning change and the degree of entrenchment in the lexicon seem to be more relevant.

As a first attempt at defining lexicalisation, following Brinton/Traugott (2005), a general distinction can be made in terms of whether the perspective adopted by the various approaches is a synchronic or a diachronic one. The former approach is mainly represented by the works of Leonard Talmy, who examines the concep-

tual representation of underlying syntactic structures. His main focus is on how a particular language encodes abstract conceptual structures and how languages may then be typologically distinguished on this basis. An influential hypothesis in this respect is Talmy's distinction between *satellite-framed* and *verb-framed* languages, which code the concepts of PATH and MOTION either separately or together (cf. Talmy 1985; 1991; 2000). This approach is synchronic in that it does not focus on a particular kind of evolution, but only on the more or less static encoding of conceptual structures. It is thus defined as a process "where a particular meaning component is found to be in regular association with a particular morpheme" (Talmy 1985, 59). For this present section concentrating, however, on linguistic development, I will not elaborate on this further. More importantly, a number of diachronic conceptions of lexicalisation will be examined shortly. Someone thinking about theories of language change in the history of linguistics may initially encounter the notion of grammaticalisation, traditionally defined as "the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status, e.g. from a derivative formant to an inflectional one" (Kuryłowicz 1965, 69). A very common example of this kind of language change is the grammaticalisation of the lexical construction *go to*, which has in many languages evolved into a grammatical means of expression of the (imminent) future (cf. Heine/Kuteva 2002, 161–163). Here, but also in other cases, as in e.g. the change of Lat. *mēns* 'reason' and Germanic *lika* 'body' to the adverbial markers of Romance *-ment(e)* and English *-ly* (cf. Detges 1998), an often free lexical element transforms into a frequently bounded grammatical one. This process is typically accompanied by a loss of semantic content, syntagmatic autonomy and an integration into a fixed paradigm (cf. e.g. Lehmann 2015). Traditionally, i.e. in its first mentions and definitions, lexicalisation has been viewed by linguists as a process going in the opposite direction, whereby a grammatical element transforms into a lexical one. Kuryłowicz (1965, 69) exemplifies this process by the development of the Lat. neuter plural morpheme *-a*, which in some Italian nouns transforms into a marker of COLLECTIVITY as in It. *mura* 'battlements' as opposed to It. *muri* 'walls' (cf. also the example of It. *diti* vs. *dita* discussed in chap. 3.2). Since the middle of the 20th century and these first mentions, the notion of lexicalisation has, however, become a much debated and diversely defined term. Lipka (1990, 95) states that "[. . .] *lexicalization*, in my view, is a *notational* term. This means that there is no single, correct definition of the term." It is not the purpose of this chapter to give a complete overview of the different approaches to lexicalisation (cf. Brinton/Traugott 2005, chap. 2), nor to treat borderline cases that may be approached as processes either of grammaticalisation or of lexicalisation (cf. e.g. Wischer 2000). Neither will I address the discussion about whether lexicalisation is parallel (cf. Brinton/Traugott 2005),

opposite (cf. Jakobson 1974, 142; Lehmann 1989) or orthogonal (cf. Himmelmann 2004) to grammaticalisation, nor whether lexicalisation as a reverse process of grammaticalisation is equal to (cf. Kuryłowicz 1965, 69), a part of (cf. Ramat 1992) or not a case of degrammaticalisation (cf. Norde 1998; van der Auwera 2002) – these issues are beyond the scope of this analysis. More relevant for this present study is the simple insight that lexicalisation traditionally always is defined solely in relation to grammaticalisation. Both processes are said to imply gradualness, unidirectionality, fusion, coalescence and demotivation as well as being presumed to be based on the principles of metaphorisation and metonymisation. In contrast to grammaticalisation, lexicalisation is said not to imply, among other purported consequences, an increase in frequency or productivity (cf. Brinton/Traugott 2005, 104–110). Given this very broad definition of lexicalisation as the creation of new lexical material, instead of the creation of new grammatical items, the following examples are generally considered cases of lexicalisation:

- Univerbation of complex syntagma into simple words: It. *forse* ‘perhaps’ < Lat. *fors sit an* ‘be it the case that’ (cf. Giacalone Ramat 1998, 122); Fr. *aujourd’hui* ‘today’ < OFr. *au jour d’hui* ‘on the day of today’ (cf. Blank 2001, 1602); Engl. *goodbye* < *God be with you* (cf. Brinton/Traugott 2005, 49–50, also for more examples)
- Compounding: Engl. *husband* < OEngl. *hūs bonda* ‘house master’ (Bauer 1983, 52); Germ. *Him-/Brombeere* ‘raspberry/blackberry’ (Blank 2001, 1600); Engl. *holyday* < MEngl. *holy day* (Kastovsky 1982, 164); Fr. *république* ‘republic’ < Lat. RES PUBLICA ‘public issue’ (Coulmas 1985, 254)
- Derivation: *companionship/membership*, derived by means of *-ship* < OEngl. **-sciepe* ‘creation, condition’; *break* > *breakage*; *divine* > *divinity* (Brinton/Traugott 2005, 35–36); Germ. *-heit* < Old High Germ. *haidus* ‘figure’ (cf. Lehmann 1989, 12–13; cf. Ramat 1992, footnote 4 contra this classification)
- Phraseologies: Engl. *to kick the bucket* (Blank 2001, 1596); Fr. *construire des châteaux en Espagne* ‘to build castles in the air’ (Blank 2001, 1600); Germ. *jemanden ins Bockshorn jagen* ‘to intimidate someone’/*durch die Lappen gehen* ‘to slip through someone’s fingers’; Fr. *être dans le coup* ‘to be on it’ (Coulmas 1985, 257)

Lexicalisation can, however, also be easily discussed without referring necessarily to grammaticalisation. It might be enough to say that, in all these examples, at some point a certain datum of linguistic material was created in discourse and this then became a fixed lexical unit in a particular the language. Adopting this definition of lexicalisation, the process implies the assumption of several stages:

As a general diachronic device lexicalisation affects all parts of the lexicon and establishes new linguistic material (either formal, semantic or both) as a language rule for a sociolinguistically defined group of speakers. It concerns, firstly, new complex words, borrowings, acronyms and delocutive formations; and secondly, the further development of already lexicalized material by semantic change, conversion, reinterpretation, formal reduction, fusion, loss of transparency, etc. (def2 and def3). The pathways of lexicalization turn out to be as multifaceted as the lexicon itself (Blank 2001, 1606).

In this reading, every new creation of lexical material that becomes a fixed part of the lexicon as well as the further development of this new creation is understood as *lexicalisation* (cf. also Brinton 2002, 74). The following sequence can then be outlined for these evolutionary steps (cf. Bauer 1983, 48; 1992, 142–143; Lipka 1990, 95–96; Matthews 1991, 100): in the beginning was an ad hoc creation, which may then be adopted by others in the speech community, a process sometimes called institutionalisation (cf. Hohenhaus 2005, 359), and ultimately fixed as an entrenched unit of the language's lexicon, i.e. as a fixed association of semantic and phonological structures (cf. Langacker 2008, 16–17; cf. also Brekle/Kastovsky 1977, 14–15). Illustrating this process by means of one of the nonce-formations discussed in chap. 5, an ad hoc construct like It. *anellatura* 'collection of rings' in example (64a) may have been created intentionally by a speaker who needed an expression for describing the rings on a fly-fishing bait. This leads to other speakers of the fly-fishing community taking over the derivate and including it in their inventory. The more people use the newly created word, the more it becomes conventional and entrenched, the more swiftly and automatically it is retrieved in use (cf. Langacker 1987a, 59). This last step is often assumed be accompanied by a simultaneous demotivation of the lexicalised item, i.e. with a loss of transparency. Blank (2001, 1598–1599), however, argues that meanings of complex words like *rattle snake* or *wheelchair* never are fully predictable and so are more than the mere sum of their parts. An illustrative example is given by German compounds on *-kuchen* 'cake': although *Himbeerkuchen* are cakes made of raspberries and *Schokoladenkuchen* those made of chocolate, *Baumkuchen* are not made of trees, but only resemble the annual rings of a tree, equally, *Marmorkuchen* are not made of marble, but their colour texture resembles marble, a *Hundekuchen* is not a cake made of dogs, but a cake for dogs and a *Blechkuchen* is not made of a baking tray, but on one (cf. Quirk 1985, 1527; Blank 2001, 1598–1599; Himmelmann 2004, 35–36). Demotivation is thus not special to the lexicalisation of complex words, but to complex words in general, lexicalised or not (cf. Fleischer/Barz 2012, 140–142). For instance, the weakly lexicalised Germ. *Tassenkuchen* 'mug cake' is baked in a cup (and not like the Engl. noun *cupcake* baked in a mould in the form of a cup), but it may also be possible that this compound denotes a cake resembling a cup or a mug. Therefore, it may be more appropriate

to say that the single constituting parts of a complex lexical unit become less salient but still remain analysable (cf. Langacker 2008, 17).

Lexicalisation may thus be understood as the coming into being and further stabilisation and change of new lexical items out of linguistic material of various kinds. In addition to this general definition of lexicalisation, Mihatsch (2006, 17–22; 2009, 81–83) discusses typical paths of lexicalisation of nouns in particular. These show various tendencies for further evolution. First, there is the tendency of noun meanings to become more and more specific. A typical example is represented by the action nouns Sp. *compra*/Fr. *achat* ‘shopping’ which over time have come to mean ‘purchase’, thus have become concrete nouns (cf. also Blank 1997, 393–394). Another typical evolution of the lexicalisation of nouns is a loss of relationality. This may be illustrated by Sp. *tío* which primarily denotes an uncle, i.e. a man who can only be defined in relation to a parent, but which has been further lexicalised to ‘guy’ in the sense of ‘that guy over there’ (cf. Mihatsch 2006, 20–21). Additional evidence for the linguistic preference of absolute noun senses over relational ones also comes from research on language acquisition. Children for instance tend to interpret a nouns like *uncle* as referring to ‘friendly man with a pipe’ and not to ‘brother of one parent’ (cf. Gentner/Boroditsky 2001, 221–222). These changes often go hand in hand with the loss of the constituting frame and thus a stabilising of the referents. Consequently, a typical path of lexicalisation of a noun begins with any kind of linguistic material, which changes first into an atypical noun with a complex gestalt, no inherent semantic properties and a high degree of relationality to then become a more typical noun with a simpler gestalt that is apprehensible holistically, has a stable extension and can be defined without any kind of context (cf. Mihatsch 2006, 21). These lexicalisation paths may be summarised as an optimising tendency towards basic level nouns, thus typically concrete count nouns having a short morpho-phonological form and referring to holistically perceivable individuals. Basic level concepts are said to generally have a higher degree of entrenchment than those of super- or subordinate levels. This correlates with a higher frequency of use as well as an earlier acquisition of these terms in child development (cf. Schmid 2010 for the relation between entrenchment and the basic level of categorisation; cf. Rosch et al. 1976 for the basic level in general). For the sake of clarification, it should be added to this short overview of the general tendencies in noun evolution that the majority of reified, non-relational nouns often do not lose their original meanings, but become polysemous. This is best exemplified by deverbal action nouns where in many cases the concrete object meaning is just added to the original semantics without displacing it (cf. e.g. Engl. *translation* ‘result/process’, Fr. *emballage* ‘process of packaging/the package itself’, Sp. *construcción* ‘process of constructing/the result of this process’). Moreover, abstract meanings are often verbalised

through the metaphorisation of concrete ones, for instance via the shift from HEAD to CENTRAL FUNCTION in many languages (cf. Blank 1998). This overview should thus not be understood as a postulation that every result of meaning change in nouns is a concrete noun on the basic level of categorisation, but simply that this is a very typical lexicalisation path of nouns.

Summarising, lexicalisation is the creation of new lexical material, its further integration in the lexicon of a language and its simultaneous entrenchment as a lexical unit. Particularly in the case of nouns, this can be viewed as a process of constant cognitive and linguistic optimisation and consequently as a tending towards the basic level of categorisation verbalised by basic level nouns. Having now elaborated the theoretical concept of lexicalisation and typical pathways of lexicalisation of nouns, I will in now focus on current research on the lexicalisation of collections.

7.2 The phylogeny of collection nouns

The last section illustrated that the coming into being of new words and phrases is governed by a number of cross-linguistically valid principles. These aspects lead to several general lexicalisation paths of which some are particular to the evolution of nouns. In what follows, the focus will lie on these latter principles and the state of the art with respect to the general evolution of collection nouns, especially in Romance languages. I will first present research done in the framework of traditional historical linguistics which has concentrated mostly on the development of collective suffixes from Latin to Romance languages. I will then focus on more recent research in the field of cognitive historical linguistics. In this respect, Mihatsch (2006; 2016) has elaborated some evolutionary principles governing superordinate nouns in general, of which collection nouns, as they are understood in the framework of this present study, are also a sub-type. The section will finally present research by Scott Grimm and Beth Levin, who concentrate on the linguistic characteristics of SOMNs, for which they particularly assume the status of artefact nouns – at least for typical representatives of this nominal category. As we shall see, this aspect may also have an influence on the diachronic evolution of collection nouns.

Before coming to the discussion of these three possible approaches, a short note on the respective focus adopted is necessary. As we will see in what follows, the different theories mostly concentrate on the second step of lexicalisation, namely the further semantic development, while assuming the preceding step of institutionalisation mainly as an evident preliminary stage which is not necessarily commented on. Although this chapter thus discusses the state of the art on

the lexicalisation of collection nouns it would be more appropriate to say that the discussion focuses on the semantic evolutions involved in this process.

7.2.1 Traditional historical linguistics

The evolution of nouns denoting collections was already thematised by Hermann Paul in his seminal work *Prinzipien der Sprachgeschichte*, first published in 1880. In his description of different types of semantic change, he identified the now well-known cognitive principles of meaning evolution being governed by, among other linguistic phenomena, metaphors and metonymies (cf. Blank 1997). For instance, he gives a series of examples of body parts being metaphorically transferred to other semantic domains like *the neck of a bottle* or *the foot of a mountain* (cf. Paul 1886, 80). Names of collections are, according to him, often derived from nouns denoting some state or characteristic. To illustrate this claim, he mentions examples like Germ. *Mannschaft* ‘team’ which goes back to Middle High German *manschaft* ‘state of being a man/relation between an overlord and his vassal’. Through metonymic change, the feature is then transferred to the collection of people being characterised by it (cf. Paul 1886, 81; GDW, s.v. *Mannschaft*). He ascribes this kind of property mainly to derivatives ending on *-schaft*, *-heit* and *-tum*, which were more or less synonymous in earlier stages of language evolution and only later drifted apart semantically (cf. Paul 1886, 215; cf. also Erben 2003, 2531; chap. 3.2). Another kind of systematic polysemy with respect to collection nouns is given by Darmesteter (1887, 62), who lists some deverbal collective nouns displaying a polysemy between an action and the metonymic result. This is the case with e.g. Fr. *ameublement* ‘action of furnishing/furnishings’ or Fr. *attroupement* ‘action of gathering/crowd’. Paul (1886, 81–82) systematises these meaning transfers by the following proposal of possible semantic changes of *nomina actionis*:

In diesen Fällen ist die Bezeichnung der Handlung auf ihr Subject übergegangen, sie kann aber auch auf das Object übergehen, Object im weitesten Sinne genommen; so auf das innere Object, wodurch eine Bezeichnung des Resultates entsteht: *Riss, Sprung, Wuchs, Zuwachs, Erhöhung, Vertiefung, Abhandlung, Versammlung, Vereinigung, Bildung*; auf das äussere Object, welches irgendwie von der Tätigkeit berührt wird: *Saat, Ernte, Spruch, Sprache, Gang, Durchgang, Übergang, Einfahrt, Zuflucht, Ausflucht, Wohnung, Kleidung*; so entstehen also auch Bezeichnungen für den Ort, wo etwas geschieht, für das Mittel, wodurch etwas bewerkstelligt wird, u. dergl.

‘In these cases, the designation of the action has passed to its subject, but it can also pass to the object, object in the broadest sense; to the inner object, which gives rise to a designation of the result: *crack, growth, increase, deepening, treatise, assembly, unification, formation*; to the outer object, which is somehow affected by the activity: *seed, harvest, saying, language*,

passage, transition, entrance, refuge, evasion, dwelling, clothing; in this way, designations are also created for the place where something happens, for the means by which something is accomplished, etc.⁶⁴

Philologists at the end of the 19th century were thus already familiar with the underlying principles governing the generation of many derived concrete nouns, namely the subject performing some action, objects involved in that action or locations where the action takes place. The driving force in these kinds of process is metonymic change. It should be noted, however, that these contributions focused mainly on semantic change and its driving forces in general and event nouns and possible polysemies in particular. Collection nouns were thus only possible instantiations of some general semantic principle and not of special interest for these works.

The same applies to a monograph published in the early 20th century. Collin (1918) analyses the Romance continuations of Lat. *-ATA* in concentrating on its primary function to derive action nouns. In this framework, he also classifies possible secondary meanings, namely the results or agents of that action, thus amongst others collection nouns. As already indicated in chap. 3.2, Lat. *-āta* goes back to *-ātus* which derives action nouns of the 4th declension. In spoken Latin, this group of nouns passes into the 2nd declension group, *viz.* nouns on *-tum*. It is very common for Latin neuter abstract nouns to be used in the plural, often without any meaning change. By means of this attraction point of analogy, the nouns formerly belonging to the 4th declension group are also used in the plural, ending on *-āta* and being finally reinterpreted as feminine singular. This was the case with e.g. Lat. *consultus, -ūs* ‘advice’ which was first transposed to the 2nd declension group, thus *consultum, -a*, was then pluralised as *consulta* and reinterpreted as feminine singular. As an illustration, Collin (1918, 43) cites an example from the *Vita Gallieni* written (supposedly) by Trebellius Pollio at the beginnings of the 5th century: “*Consulta Valeriani fratris sui*” (‘on the advice of his brother Valerianus’), where the noun in question can only be feminine singular (here in the ablative case). Since a detailed description of these formal evolutions has already been given in chap. 3.2, I will not comment further after this short recap. More important are the semantic changes Collin describes by considering many examples from various Indo-European languages, he develops a classification of possible meaning extensions of action nouns:

⁶⁴ Some of the translated English examples have two equivalents in German, which differ slightly in meaning (e.g. Germ. *Gang/Durchgang* ‘passage’). For this reason, the number of examples given differs between the German original and the English translation.

- The agent of the action, e.g. Swed. *bevärning* ‘armament’ > ‘soldier’ > ‘soldiers (coll.)’ (cf. Collin 1918, 57–59); other examples specifically derived from Lat. -ĀTA: Fr. *volée* ‘action of flying > swarm = birds which fly together’; Fr. *cavalcade* ‘ride > cavalcade’ (cf. Collin 1918, 176–177)
- The object of the action (in reference to Paul (1886) he also differs between an inner and an outer object), e.g. Lat. *lectio* ‘reading > reading material’ (> Fr. *leçon* ‘lesson’) (cf. Collin 1918, 59–64); other examples specifically derived from Lat. -ĀTA: Fr. *fumée* ‘action of smoking > smoke’; Sp. *hallada*/Pt. *achada* ‘action of finding > find’ (cf. Collin 1918, 177–211)
- The instrument of the action, e.g. Lat. *vectura* ‘transport’ > Fr. *voiture* ‘carriage’ (cf. Collin 1918, 64–65); Collin does not cite any further examples of instruments derived by Lat. -ĀTA
- A designation of a time span or a point in time in relation to the action, e.g. Fr. *au lever du soleil* ‘by sunrise’ (cf. Collin 1918, 65–66); other examples specifically derived from Lat. -ĀTA: Fr. *journée* ‘day’, Sp. *temporada* ‘period of time’, It. *invernata* ‘period of winter’ (cf. Collin 1918, 212–219)⁶⁵
- A designation of a place where the action takes place, e.g. Anc. Gr. ἀγορά ‘public reunion’ > ‘assembly/gathering place’ (cf. Collin 1918, 65–69); other examples specifically derived from Lat. -ĀTA: Fr. *baignade* ‘action of bathing’ > ‘bathing area’; It. *entrata* ‘action of entering’ > ‘doorway’ (cf. Collin 1918, 220–225)

The first two categories are of particular interest for delineating the evolutionary paths of collections since, regarding the agents of an action, Collin (1918, 57–58) notes that

le plus souvent, [le nom d'action] prend alors un *sens collectif*. [. . .] Il peut aussi passer du sens collectif au *sens individuel*. Ainsi, *recrue* désigne d'abord l'ensemble des soldats nouvellement enrôlés, puis un seul de ces soldats. (emphasis in the original)

‘more often than not, [the action noun] then takes on a collective meaning. [. . .] It can also pass from a collective to an individual meaning. For example, Fr. *recrue* first refers to all newly enlisted soldiers and then to a single soldier.’

With respect to the objects of an action, he does not explicitly generalise a collectivising potential of these result nouns, but he states e.g. that “Lat. *piscatum* avait déjà en latin pris le sens collectif de poissons pris en une seule pêche” (“Lat. *pis-*

⁶⁵ Note that Collin treats examples like *lever* and *journée* alike in considering them as instantiations of the meaning evolution of an action noun to a time designating noun. Whereas there is indeed an action amid the origins of *lever*, one cannot detect such a meaning for *journée*. We must consequently suppose that the suffix *-ée* already inherited the capacity of deriving time span denoting nouns.

catum had already taken the collective meaning of fish taken in only single catch') (Collin 1918, 60). One can thus deduce from his findings a typical path of evolution of action nouns to designate either the subject or an object involved in that action. These subjects or objects may often be involved in that action at a number higher than one so that a collective meaning may evolve. From this collective meaning the sense of an individual may evolve. He explains these meaning transitions following the Danish linguist Sandfeld Jensen with *forvekslingsmuligheder* 'possibilities of confusion'. The confusion lies in the formal resemblance between abstract and concrete nouns (in Indo-European languages, there is no morphological indicator differentiating them), in the same possible syntactic positions they can take and finally, in the often polysemous verbs they can be combined with. Verbs like *to take* may have an abstract or a concrete meaning, initiating the transition of meaning in the noun following it (e.g. *take an apple* vs. *take sb. in marriage*) (cf. Collin 1918, 116–118).

In referring to the work of Paul (1886) and Collin (1918), Baldinger (1950) structures his description of French collective suffixes and derived nouns formed by means of them. A collective meaning may then arise from an action in either referring to the collective subject of that action or to a collective object (e.g. the result or an instrument). Since his work has already laid the theoretical foundation for my own considerations and empirical analyses in chaps. 3.2 and 5, I will in the following only stress the most important aspects of his work, enriching it with the results of my research on collections as a morphological category. As already mentioned, Baldinger distinguishes between spontaneous and functional collections, the former being deverbal nouns whose collective meaning only came into being via semantic change, the latter being denominal nouns derived by means of a collective suffix. Baldinger (1950, 110) emphasises that the capacity for semantic transfer from abstract event nouns to concrete collective nouns is not a typical property of collective suffixes, but of *nomina actionis* in general. He illustrates this by means of Fr. *parachutage* which is first attested in the first half of the 20th century and at that time meant either the action of airdropping or its collective result, the people or goods airdropped. Note, as has also been shown by my analysis of nonce-formations, that Fr. *-age* synchronically has no direct collectivising potential, "nur noch eine spontane, deverbale kollektive Kraft" ('only a spontaneous, deverbal collectivising force') (Baldinger 1950, 31). A number of factors influences the exact outcome of these processes of semantic change. For instance, the semantic domain of the verb involved in the particular lexicalisation process may determine the possible collection. It is thus obvious that only transitive verbs may lead to a deverbal collective noun denoting some result of an action, like Fr. *rendement* 'yield' (< *rendre* qqc. 'to yield sth.') (cf. Baldinger 1950, 113) and that only in combination with the matching suffix can a collective deri-

vate be the outcome of the derivational process. For instance, as has been shown in chap. 5, the derivates of Fr. *-erie*, Sp. *-ería* and It. *-eria* are especially prone to the semantic domain of tools, crafting (usually by hand) and its result, so only in combination with derivational bases of the same semantic domain may they lead to a collection denoting some kind of handicraft. In accordance with the default inheritance principle, these restrictions on possible outcomes are governed by the superordinate constructional schema, valid for nonce-formations as well as for lexicalised derivates. The examples analysed in chap. 5, however, also made clear that the boundaries of these constructional schemata may be transcended, i.e. that the implications and restrictions of a schema are altered. This was the case of Fr. *sachetterie* ‘collection of bags/action of bagging up’ in example (56e), where a denominal derivate shows the same event/collection polysemy as the deverbal equivalents, although there is no event-related derivational base. Functional collectivity is, in contrast, found in all those cases where a collective suffix directly bounds all the underlying referents of the base noun. Formally, this is manifested by denominal derivations as opposed to deverbal derivations in the case of spontaneous collections. Collective suffixes deriving these kinds of collections might already have some meaning of plurality in Latin, e.g. Fr. *-aille*, or they developed this function by metonymic change, e.g. Fr. *-ment*.

Baldinger’s analysis of French collective suffixes, also summarising research done by Paul (1886) and Collin (1918) amongst others, thus leads to the conclusion that there are three possible pathways of lexicalisation of derived collections: the metonymic change from events to collections as well as from properties to collections, and the lexicalisation of the Latin neuter plural (cf. Baldinger 1950, 215–216). These paths can be illustrated by the semantic map of COLLECTIVITY in the Romance languages represented in Figure 5.9. The map is repeated here for reasons of convenience as Figure 7.1, though the additional semantic extensions are not shown:

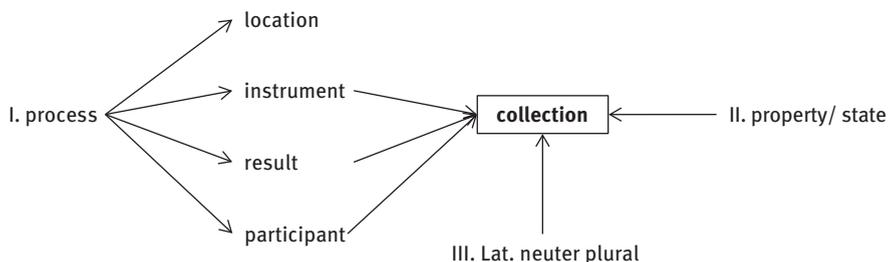


Figure 7.1: Lexicalisation paths of derived collection nouns in Romance.

In accordance with lexicalisation theory established in chap. 7.1, pathway I and II represent typical cases of metonymic concretisation as argued by Moreno Cabrera (1998), while pathway III represents a typical example of inflectional morphemes becoming derivational, and thus becoming more lexical.

Research undertaken in the realm of traditional historical linguistics has already indicated typical origins of collections and resulting evolutionary paths. The focus, however, was always on the starting and end points of linguistic development, while possible intermediate stages were of lesser interest. Neither have different kinds of collection, e.g. with respect to their countability, been analysed separately. These issues are now addressed in the framework of cognitive historical linguistics, mostly represented by the work of Mihatsch (2006; 2016).

7.2.2 Cognitive historical linguistics

As we have seen in the last chapter, traditional historical semantic approaches mostly adopted the methodology of semasiologically analysing various derivatives to then onomasiologically determine meaning relations and meaning changes, i.e. for instance, the metonymic change from the designation of an event to the denotation of the objects involved in it. Since the 1980ies, another approach combining historical linguistics and the newly emerged branch of cognitive linguistics was developed. The work of i.a. Geeraerts (1983) or Traugott (1985) laid the basis for the present understanding of cognitive approaches to diachronic semantics. Strictly speaking, these theories picked up the thread of the pre-structuralist approaches discussed in the last section. What is new to this perspective is the additional step of combining the observed regularities in language change with human cognitive processing. They detect regularities in language change and connect it to human cognition like e.g. in the case of conceptual metaphors (*ANGER is HEAT*). In Romance linguistics, it was especially the work of Peter Koch and Andreas Blank amongst others which played a pioneering role in establishing a cognitive understanding of particularly historical onomasiology (cf. i.a. Blank/Koch 2003b; see in particular the introduction for a comprehensive overview). They add the level of extra-linguistics concepts, especially the domain of cognitively very salient body parts to the language-specific onomasiological level. Under this perspective, semantic change is not only detected and described (as has been done in the traditional approaches), but explained via principles of human thinking and perception (like e.g. effects of figure-ground-change).

The research of Mihatsch (2006; 2016) has to be understood in this framework. She does not concentrate specifically on nouns denoting collections, but on superordinates in general. For these kinds of noun, she draws a unidirectional

lexicalisation path, along which different types of collection nouns are possible intermediate steps. Many hyperonyms denoting individual superordinate categories originate in so-called functional nouns. These are weakly lexicalised nouns like Fr. *achat* ‘purchase’ which have no stable intension nor extension. Consequently, these functional nouns often denote ad hoc categories like *things to take on a camping trip*, for which only the contextual frame determines the actual referents; what exactly is purchased is not specified by the meaning of the noun but only by the context. Functional nouns often comprise more than one referent, e.g. Fr. *couvert* ‘things with which one covers the table’ > ‘cutlery’ (cf. Mihatsch 2006, 115–116; cf. also Barsalou 1983). Because of their possible referential plurality, many functional nouns develop during lexicalisation into CCNs. These are still highly dependent on the contextual frame and thus represent heterogeneous sets. A typical example for this kind of collection is Engl. *outfit*. There is possibly a class of referents that may typically constitute an outfit, but generally these kinds of collection are “linked to particular events or situations and are spatio-temporally restricted” (Mihatsch 2016, 294). Cruse (1986, 99) calls these “relations[s] defined in terms of expectation *para-relations*” (cf. also Mihatsch 2006, 117–118). Other linguistic features match what has been elaborated for typical CCNs in chap. 3.1 and summarised in Figure 3.1: they are countable (*two outfits*), the relation between the collection and its members is meronymic and not hyponymic (**outfit industry*) and the constituting entities are hardly accessible (*?tight-fitting wardrobe*) (cf. Mihatsch 2016, 294–295). Examples of this kind are particularly common for the semantic domains of CLOTHING and FURNISHINGS, for instance: Fr. *vestiaire/garde-robe*/Sp. *ropero* ‘wardrobe’ > ‘clothing of a person’; Fr. *costume* ‘dressing style of a particular social group’ > ‘clothing of a person’ > ‘suit’; Sp. *atuendo* ‘outfit’ (< Lat. *ATTONITUS* ‘amazed’); Sp. *mobiliario*/Fr. *mobilier* ‘movable (property)’ > ‘furniture’ (cf. Mihatsch 2006, 118–121). All these examples mostly originate in functional nouns and are still characterised by a dominant frame constituting the actual reference. Koch (2005) calls these contiguous relations between conceptual frames and their elements *engynomic* to overcome the traditionally rather narrowly defined term of contiguity (cf. also chap. 3.1.2). The extension of these kinds of collection noun is more stable than that of their original functional nouns in being characterised by logical expectations. The referents, however, remain a heterogeneous set. Progressing further in the lexicalisation process, some of these collective nouns may lose their frame and become superordinate OMNs. This process is accompanied by a change from collections being constituted by the principle of contiguity to collections being characterised by the principle of similarity: “C’est en passant du collectif-ensemble [i.e. CCNs] au collectif-genre [i.e. OMNs] que l’engynomie se transforme en taxinomie” (‘it is by passing from a set-collective [i.e. CCNs] to a category-collective [i.e. OMNs]

that engynomy becomes taxonomy’) (Koch 2005, 185; cf. also Mihatsch 2006, 123–124). Since they are spatio-temporally unbounded, OMNs like *clothing* may have a more stable and homogeneous set of constituting entities than collective nouns like *outfit*. As Mihatsch (2006, 125; 2016, 296) illustrates: Sp. *ropa* ‘clothing’ typically only comprises outer clothing, without hats, shoes and accessories, and also Engl. *furniture* comprised in its earlier stages all kinds of furnishings including fabrics and decorations, whereas the OMN nowadays, however, typically only designates bigger pieces of wooden furniture (cf. Mihatsch 2006, 126; cf. also OED, s.v. *furniture*). Summarising, “die Eigenschaften, auf denen Genuskollektiva basieren, sind stabiler, inhärenter, damit intensional reicher, kontextautonom und absoluter als die der Gruppenkollektiva.” (‘the properties on which OMNs are based are more stable, more inherent and thus richer in their intension, more context-autonomous and more absolute than those of spatio-temporally bounded count collectives’) (Mihatsch 2006, 127). In addition to the factors influencing ontological homogeneity of referents as elaborated in chap. 3.1.2, one may consequently assume that collections whose constituting members are more homogeneous at the same time are older and display a higher degree of lexicalisation than collections whose members are more heterogeneous. If the level of similarity reaches a point when the constituting members of a collection may indeed be classified as one homogeneous category, the SOMN may create a morphological plural and it then becomes a POMN without an equivalent singular (cf. Mihatsch 2006, 128). Note, however, that POMNs do not necessarily need to have a developmental pre-stage of having been a SOMN, but often come into being as the fossilised inflectional plural of an object noun. For instance, the German plurale tantum *Klamotten* ‘clothes’ originated in Berlin slang *Klamotte* ‘old broken thing’ and thus has no SOMN as a pre-stage (cf. also chap. 3.2). In turn, the French SOMN *vaisselle* and its Spanish equivalent *vajilla* originate in Lat. *VASCELLA* (n. pl. of *vascellum* ‘recipient’) and supposedly have no CCN pre-stage. My impression is that referents of POMNs are not necessarily always more homogeneous than those of a SOMN: Germ. *Klamotten*_{POMN} ‘clothes’ and *Kleidung*_{SOMN} ‘clothing’ comprise both more or less heterogeneous items – when asked what to buy when the children are growing out of their *Kleidung* or *Klamotten*, one would equally enumerate outer clothing, but also shoes or accessories like scarves or woolly hats. This is, however, only a point-by-point impression and, to my knowledge, there has yet been no systematic examination of this topic.⁶⁶ Many OMNs diachronically fluc-

⁶⁶ According to Wierzbicka (1985, 282–283) the difference between Engl. *clothing* and *clothes*, similar to Germ. *Kleidung* and *Klamotten*, lies in the fact that the former refers to pieces of clothing not necessarily fitting together and the latter to a functional set which is bounded to a place and a moment of time.

tuates between the morphological singular and plural, this is e.g. the case with Sp. *fruta(s)* ‘fruit’, Fr. *sape(s)* ‘clothing/clothes’ and also Engl. *vegetation(s)* and *furniture(s)* (cf. Mihatsch 2006, 129). If the plural then becomes dominant, there is the possibility of an inflectional plural, resulting in a countable object noun at the superordinate level (cf. Mihatsch 2006, 133–135).

The path traced from functional noun, to CCN, to SOMN and POMN, to finally result in a countable object noun can be exemplified by means of Sp. *ropa* ‘clothing’ (cf. Mihatsch 2016). This goes back to the Gothic verb RAUPJAN ‘to rob’ which in turn forms the result noun **raupa* ‘booty/loot’ (cf. DECH, s.v. *robar*). Consequently, the origin of *ropa* is a typical functional noun with the meaning ‘what has been robbed’. This highly context dependent interpretation as a result noun is not attested for Spanish, but in the OHG *Hildebrandslied*, we find *wala-raupa* ‘the looting of the fallen warriors of the Walstatt’ (FEW, s.v. *rauþa*). The Germanic result noun then lexicalises further into a CCN in Old Spanish having the meaning ‘robbed equipment or clothing’, which is still spatio-temporally bounded, but now the extension is limited to a specific referent class, i.e. by para-relations (cf. Mihatsch 2006, 118). Because of the gradual fading of the frame, the referents have become more homogeneous and more stable, and *ropa* is now limited only to clothing excluding shoes and accessories (cf. Mihatsch 2006, 125). The morphological plural of *ropa* seems to always be present, as well as *ropa* meaning ‘piece of clothing’. It may thus be assumed that, once established, the collection oscillates between its different forms and the stages through which it has passed. This may be determined and restricted by various factors, like conflicting forms (cf. Mihatsch 2016, 303).

One of the possible pathways of lexicalisation of hyperonyms, of which different types of collection nouns are possible intermediate steps, may be summarised as follows:

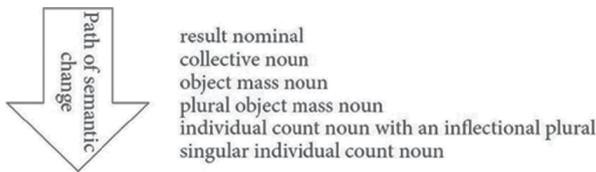


Figure 7.2: Assumed pathway of lexicalisation of collections (Mihatsch 2016, 301).

I will now turn from this description of Mihatsch’s theory, as further questions and hypotheses arising from her assumptions will be addressed in chap. 8.

7.2.3 Artefacts as SOMNs

Whereas the concept of lexicalisation as well as the description of assumed paths of lexicalisation were both discussed adopting a historical linguistic approach, this final part of the presentation of the state of the art will focus on synchronic research effectuated in the framework of formal semantics. As will be shown in what follows, this approach adds an interesting piece to the puzzle of the evolutionary paths of collection nouns. Specifically, Scott Grimm and Beth Levin have carried out a series of tests examining the linguistic properties of English SOMNs in present-day use, or *furniture-nouns* in their terminology (cf. Grimm/Levin 2011; 2012; 2016; 2017). In this respect, they are especially interested in research questions like these nouns' uncountability, of which kind of entities they are comprised and the degree to which these are linguistically accessible. One of their main assumptions is that SOMNs are artefact nouns contrasting with nouns denoting natural kinds. Whereas natural kinds are constituted by properties referring to their outer appearance, their surroundings and so on, artefacts are determined by an event associated with them. On the basis of this argument, they differentiate between functional artefacts, whose associated event is only potential, and stage-level artefacts, whose associated event has actually happened. Examples of nouns denoting functional artefacts are *cup* and *hammer*, while *delivery* or *gift* are examples of stage-level artefacts. The authors state that SOMNs like Engl. *furniture*, *mail* or *change* are artefact nouns, since they are always linked to a certain event. *Furniture* is classified as a functional noun and *mail* or *change* as stage-level artefact nouns. One piece of evidence they cite for this classification is the fact that most SOMNs originate as either deverbal nouns (*furniture* < Fr. *fourniture* 'equipment' < Fr. *fournir* 'to furnish/to equip') or are in their origins otherwise closely linked to a certain event (*mail* < *mail of letters* (existing only in one particular event of mail delivery)). Their conclusion with respect to countability and other linguistic properties following from these considerations are not of further importance for this present chapter (cf. chap. 3.1), but what is relevant is that they classify SOMNs as artefact nouns in contrast with natural kinds. In this respect, they implicitly raise the crucial question of whether the regularities in the evolution of collection nouns as assumed by the tenets of historical linguistics, are valid for all these kinds of noun, or whether the mostly metonymic changes described can only be found in the case of artefactual collection nouns. Indeed, as shown in Table 4.2, there are e.g. much fewer animate and human OMNs than inanimate ones. Animate OMNs like Fr. *bétail* 'cattle', Fr. *volaille* 'poultry' or human ones like Fr. *gens* 'people' furthermore show a greater diversity in their origins, whereby they are not necessarily based on functional

nouns, deverbal derivation or the like. The empirical analysis in chap. 8 will need to shed further light on these issues.

Having considered the state of the art on the lexicalisation of collection nouns, the next chapter will briefly outline additional evidence for the regularities summarised above in considering the ontogeny of collection nouns.

7.3 The ontogeny of collection nouns

In the framework of this present thesis, I assume a very basic parallel between the diachronic evolution of a language and language acquisition and evolution in child development. As a consequence, the examination of collection nouns in language acquisition may add another piece to the puzzle of, and an additional piece of evidence for, regularities in the evolution of collection nouns in diachrony. In this respect, Diessel (2012) assumes three main parallel principles in language change and language acquisition based on similarities in language use, namely analogy, entrenchment and categorisation. Analogy is a well-known driving force particularly for morphological change and can often be found in e.g. the tendency to regularise irregular paradigms. For instance, many irregular forms of the past tense in English were regularised from Old to Modern English (e.g. *low* > *laughed*). In a parallel way, children acquiring English first overgeneralise the regular pattern and produce forms like *blowed*, and only later they do recognise the irregular form and produce it correctly. These tendencies to regularise can be found in phylogenetic language evolution, e.g. from Old English to Modern English, as well as in ontogenetic child evolution, i.e. from one stage of language acquisition to another (cf. Diessel 2012, 1603–1605). Furthermore, linguistic entities are more immune to regularisation processes the more they are entrenched in the memory of a speech community or of a child. Entrenchment is facilitated when a linguistic entity occurs more often in language use. For this reason, many highly frequent verbs have irregular inflection patterns, because they were able to resist regularisation due to the common use of their original, irregular forms (cf. Diessel 2012, 1604; cf. also Diessel 2007, 117–119; cf. Langacker 1987a; 2017 for entrenchment in general). Finally, categorisation plays a role both in language evolution, particularly grammaticalisation, and language acquisition, since both processes mostly rely on the same principles of transfer of conceptual domains (e.g. *SPACE* > *TIME*; cf. Diessel 2012, 1605–1607; cf. also chap. 7.1). These examples show that language change and language evolution share a number of principles, which are, however, mostly restricted to cognitive and conceptual mechanisms (cf. Diessel 2011, 138–140). Although I am not postulating an absolute parallelism between the phylogeny and the ontogeny of language

(cf. i.a. Traugott/Dasher 2002, 42–44), these very general principles underlying and governing the ways of think and speak are especially relevant for the present analysis.⁶⁷ Before turning to the acquisition of collection nouns, a brief remark on noun acquisition in general is necessary. In chap. 7.1, I described the paths of lexicalisation of nouns which, summarising the different aspects, go from abstract to concrete, from complex to simple and from relational to absolute features. In a certain way, the acquisition of nouns in child development goes in the opposite direction, reflecting nevertheless the tendencies of nominal evolution: whereas nouns in diachronic evolution tend to take on structures and features that are cognitively easier to handle, children begin with these cognitively simpler structures and only later change to more abstract and complex ones. These cognitively simpler words are mostly words for directly perceptible and holistically apprehensible referents, words that refer to the child's immediate environment (cf. Wode 1988, 144–147). Following on from this, nouns are typically acquired earlier than verbs, pronouns and other linguistic categories that tend more to depend on contextual information and cognitively more complex processes (cf. Gentner/Boroditsky 2001). Within the word class of nouns, children typically acquire proper names of human entities as well as for inanimate concrete objects first, reflecting the ease of acquisition of highly individualised, directly perceivable and holistically simple entities (cf. Bassano 2005).

Categorisation based on different kinds of cognitive principle seems to play a role in the evolution of collections in diachrony, but also in child development. Specifically, experiments analysing the way children categorise collections of extra-linguistic entities seems to indicate parallels between the constitution of collection nouns in language evolution. The description of the state of the art on the evolution of collection nouns in chap. 7.2 has shown that many collection nouns pass from being CCNs characterised by meronymy to OMNs being constituted by both meronymy and hyponymy (cf. chap. 3.1.2). Indeed, various experiments with small children show that their primary categorisation of extra-linguistic entities relies on meronymic part-whole relations – hyponymic class inclusion as a means of categorisation is not acquired until later in children's development. Markman/Seibert (1976, 565) summarise this path of development on the basis of various works from the 1960s (including those by e.g. Jean Piaget)

⁶⁷ It should be noted that the basis for this linguistic theory, the biogenetic law or recapitulation theory, neither is undisputed. Going back to E. Haeckel, it assumes that the ontogeny of a living being recapitulates its phylogeny. This is, however, only partly true and should not be generalised to such an extent that it is applied as a universal rule. Nevertheless, the theory may, for instance, be helpful in explaining deviant evolutions in the ontogeny as historical residues of the phylogeny. such as the gills anlage in tetrapods (cf. LB, s.v. *Biogenetische Grundregel*).

as follows: “In sum, when instructed to ‘put things together that go together or are alike’, children form classifications that seem to move from objects to collections to classes.” The authors tested this classification continuum over three experiments by showing children of different ages (kindergarten and primary school) various objects either labelling them as members of a collection or as a class (e.g. baby frogs and their parents as members of a frog family vs. the same referents as exemplars for the frog class). The children then had to make a quantity judgement depending on the test condition. They were either asked who would have more pets, choosing between the two options of someone who owned the baby frogs or someone who owned the family (collection condition), or choosing between someone who owned the baby frogs or someone who owned the frogs (class condition) (cf. Markman/Seibert 1976, Exp. I). In all test conditions the children found it easiest to make these quantity judgments when a CCN like *family* was named together with its constituting entities, whereas all the children had problems with conceptualising the class inclusion questions (cf. also Callanan/Markman 1982). The authors explain these differences by the fact that individual objects and contiguous collections like *family* can be perceived holistically, whereas a comparison of quantity is much more difficult for a class that cannot be perceived as a whole (cf. Markman/Seibert 1976, 574; cf. also Markman/Horton/McLanahan 1980 for similar results on novel categories). They also tested the hypothesis that the grammatical number of the target noun may play a role. Since the CCNs were always presented in the singular (*family*) and the count superordinate nouns in the plural (*frogs*), one may assume that the singular favours the interpretation as a holistically perceivable whole. To test this assumption, in a second experiment they included SOMNs (*food, money, furniture* and *silverware*), with the methodology being identical to the first experiment (cf. Markman/Seibert 1976, 571–572). The results show that children had similar problems making these quantity judgements (e.g. “who would have more to eat, somebody who ate the cookies or somebody who ate the food?”) as they had with the count superordinates – the inflectional number thus does not seem to play a role here. In sum, the results show that it is difficult for children to make quantity judgements for class inclusion relations since they have to keep in mind a whole class which however is not perceivable as a whole. This implies that nouns for holistically perceivable collections are acquired earlier than classes – a generalisation that parallels the diachronic evolution of collection nouns assumed by Mihatsch (2006; 2016).

In addition, Markman (1985) more specifically investigated the question of whether there is a difference between countable and uncountable superordinate terms like *vehicle* and *furniture* by testing the hypothesis that the latter may be a kind of compromise between prototypical classes and collections. In her second study, which is most relevant here, she taught children of kindergarten age new

categories labelling them either with a fantasy count or a mass superordinate. For instance, children were shown different bathroom supplies. In the count condition this new category was labelled *These are vebs* and in the mass condition *This is veb*. After the introduction of this new category, children had to decide whether new items belonged to that category or not (cf. Markman 1985, 43–46). The results show that the generally difficult task was facilitated for the children when they learned the new category when it was labelled with a mass noun. The author thus concludes that mass superordinates (i.e. SOMNs) do indeed represent a kind of compromise between collectives, which are cognitively more simple but cannot serve for categorisation, and countable hyperonyms, which are cognitively more difficult but necessary for categorisation: “Thus mass nouns can encode superordinate categories in a way that simplifies the hierarchical representation yet remains faithful to the inclusion relation” (Markman 1985, 51).

More evidence for this primarily contiguous interpretation of extra-linguistic objects comes from Macnamara (1982). In a series of verbal and non-verbal categorisation experiments with 2½-year-olds, he found on the one hand that they refuse to label an individual object like a horse with a superordinate term like *animal*, but on the other hand they are well able to interpret such a term when applied to individuals. For instance, the children were asked questions like “Is this an animal?” while being shown individual animals like a horse or a dog. Most children vehemently responded to these kinds of question with statements like “No, piggie” (cf. Macnamara 1982, 63–67). When, however, they were asked to take either the animal or the toy held by the tester, they were generally able to correctly apply the noun to the referent and take the requested object (cf. Macnamara 1982, 76–78). The author concludes from this that children at the age of 2½-years “treat [even count] superordinate terms like collectives [. . .], [. . .] [since they] become acquainted first with the terms *toy* and *animal* as used of assortments of objects” (Macnamara 1982, 81). This therefore not an explanation for the conceptual basis of categorisation, but rather a usage-based account based just on the linguistic cues that children receive and on which they rely.

Summarising the various studies on the acquisition of count and mass superordinates, children first acquire names for individual objects on a subordinate level. When confronted with some kind of superordinate concept, they first interpret it by means of a holistic collection where the individuals are related to the superordinate by the principle of meronymy. Even count superordinates like *animal* are first interpreted by them as bounded collections and not as names for categories. Only later do they begin to change from meronymy to hyponymy, from collections to categories. Mass superordinates like *food* or *furniture* may help them achieve with this cognitive categorisation transition, since these are on the boundary between meronymy and hyponymy.

One can thus assume that mass superordinates may not only facilitate categorisation, but are also acquired earlier than count superordinates. This would in general imply an acquisition continuum from holistically perceivable objects labelled first by object nouns, then by CCNs, then by SOMNs and finally by count superordinates. These ontogenetic findings would then reflect the lexicalisation path of collections assumed here.

8 Lexicalisation of collection nouns: Corpus analysis in *Frantext*

The next part of this diachronic section will focus on empirically examining the theoretical issues raised in chap. 7. In the following preliminaries, I will first explain my methodological choices and delimitations (chap. 8.1). The empirical analysis then addresses the etymology of French collection nouns focussing on the starting point of a possible lexicalisation path by means of a lexicographic analysis (chap. 8.2), as well as the further development of these kinds of noun by means of a corpus analysis in the *Frantext* corpus (chap. 8.3).

8.1 Methodology

The following corpus analysis aims to elaborate the theoretical path of lexicalisation of collection nouns on an empirical basis. As previously pointed out, there are no systematic, exhaustive corpus analyses that actually prove that there is such a path for Romance collection nouns. The empirical basis consequently consists in the systematic tracing of the diachronic evolution of a number of collection nouns. Before coming to the description of the actual choice of nouns and the exact research questions tied to it, a number of methodological issues must be discussed. First of all, there is the language choice to consider. Since this present work does not focus on language history, but rather aims to give a fuller picture of the expression of COLLECTIVITY in French and in Romance languages in general, a comparative analysis as was undertaken in the synchronic part is not possible. This is mainly because an exhaustive corpus analysis, which is both qualitative and quantitative, is very time-consuming. In addition, corpus research in the nominal domain in particular has to be on specific nouns and not on more abstract (syntactic) constructions. More than one noun needs to be considered for generalisations to be adduced about a whole noun type. One also has to consider a long time-span of evolution to fully understand a noun's overall development. However, even for a noun that is used with a lower frequency in present-day language, like many of the OMNs assessed in chap. 4, the number of occurrences quickly rises when considering the nouns' evolution from their beginnings until now. Even taking into consideration a sample size for all occurrences in only one century, a cross-linguistic comparison would only be possible in a study focusing exclusively on this diachronic part. For these reasons, I have chosen to focus again on French. As shown by the synchronic analyses of present-day language, collection nouns in French display the most clear-cut categorical differences – when compared to the other Romance languages exam-

ined in this study. Here, we do find prototypical instances of uncountable SOMNs as well as POMNs (e.g. *habillement* ‘clothing’, *gens* ‘people’) in opposition to countable spatio-temporally bounded collective nouns (e.g. *équipe* ‘team’). The acceptability judgement tests presented in chap. 4 revealed a greater degree of flexibility of collections in the case of the other Romance languages studied, especially with respect to the OMNs analysed. I even doubted the very existence of OMNs in e.g. Portuguese given the results of this study (cf. chap. 4.2.3). Considering the assumed path of lexicalisation of collection nouns as hypothesised by Mihatsch (2016) with its different stages represented by various aspectual types, I assume French to display possible developmental differences the most clearly. It would, however, be fruitful for future research to complete this analysis by considering other Romance languages, too (cf. Mihatsch/Kleineberg 2022 for a parallel study on Spanish OMNs).

In addition to these reasons based on the results of the examination of present-day language, research on the diachronic evolution of collections has focused predominantly on the French language (cf. chap. 7.2). In this respect, especially the detailed work of Baldinger (1950), and to a lesser extent also that of Collin (1918) and Roedinger (1904), provides us with an extensive list of (derived) collections that may serve as points of comparison. In particular, those instances that are not used nowadays may be of interest for this present corpus analysis. There are no works on Italian, Spanish or Portuguese comparable to this work on French.

Finally, and this last reason has to be considered not as decisive, but merely as additional, diachronic corpus analyses of French have been made much easier by the existence of the *Frantext* corpus. There are indeed diachronic corpora for each of the ‘major’ national Romance languages (e.g. the *Corpus del diccionario histórico de la lengua Española* (CDH) for Spanish), but only *Frantext* makes available a range of convenient tools like random sample selection, efficient data download in .csv-format and the possibility to make lemma queries in medieval spelling. Although the modern corpora of the Spanish *Real Academia* or the *Corpus del Español* of Mark Davies (by now) also allow for complex queries, the handling of large data quantities is still time consuming. Consequently, the *Frantext* corpus had to be rejected for synchronic research on OMNs and collective nonce-formations, but it is more than suited for a diachronic analysis. It comprises texts ranging from the very beginnings of French language documentation until modern language use focusing mainly on literary texts.

In the following, I will first elaborate my research questions for this diachronic part based on the state of the art summarised in chap. 7 to afterwards explain the exact choice of nouns and characteristics to be examined.

The state of the art on pathways of evolution of collection nouns may be summarised as follows: artefactual collection nouns come into being as functional nouns strongly related to a particular event. For this reason, they are often dever-

bal nouns representing a systematic polysemy between actions and their related results, affected objects, instruments or agents. In their beginnings, these functional nouns have no stable extensions and their referents are highly context-dependent. Illustrative examples of this category are represented by the collective nonce-formations analysed in chap. 5. During their further development, the extension begins to stabilise and a spatio-temporally bounded CCN evolves. The referents of these nouns are still highly heterogeneous, but their extension is now restricted by para-relations, i.e. expectation relations. By further evolving, the referents become more homogeneous and the extension gets more stable. While losing its contextual bounding, the CCN changes into a non-countable SOMN which allows categorisation. Finally, the single individuals may become more and more salient. This development is reflected by the change from a SOMN to a POMN to a count superordinate. Morphologically, the pathway of lexicalisation assumes a change of state from CCN, to uncountable SOMN, to a quasi-countable POMN to finally become a count noun again. This state of the art raises the main research questions to be addressed: To what extent can this pathway of lexicalisation be proven by corpus data? The main hypothesis in this respect is that it can indeed be proven for derived artefactual collection nouns, but with respect to other morphological and ontological types of collection noun, there are two additional subordinate research questions:

RQ_{dia}1: Given the fact that the pathway of lexicalisation assumes a tight relation to events, objects and persons associated with it, do collection nouns denoting animals also follow the steps described?

RQ_{dia}2: Given the fact that we indeed predominantly have illustrative examples of deverbal nouns like Fr. *habillement* ‘clothing’ or Engl. *furniture*, to what extent have collection nouns going back to e.g. the Latin neuter plural *-ĀLIA* taken part in this evolution?

In light of the results of the synchronic analyses as well as the state of the art on the diachronic evolution of collection nouns, the following hypotheses to these research questions may be formulated:

HP_{dia}1: Given the results of the synchronic analyses of present-day language, I assume that collection nouns denoting animals do not represent prototypical representatives of this nominal category – the special behaviour of the animate OMNs in the acceptability judgement tests (cf. chap. 4) as well as the fact that there were only marginally animate nonce-formations (cf. chap. 5) indicate this. It may thus be assumed that, since animals do not represent typical agents or

objects involved in actions, their path of evolution either is not equal to, or only by analogy, that of artefactual and human collection nouns.

HP_{dia2}: Similar to the considerations made for HP_{dia1}, I also assume collection nouns which are derived with *-ĀLIA* to participate in that path of lexicalisation only by analogy (if they show the assumed steps of evolution at all).

To address these research questions and hypotheses, I analyse in what follows the diachronic development of the French OMNs examined in the acceptability judgement study, *viz.* *fringues* ‘clothes’, *bétail* ‘cattle’, *gens* ‘people’, *mercerie* ‘haberdashery’, *mobilier* ‘furniture’, *vaisselle* ‘crocery/tableware’, and *volaille* ‘poultry’.⁶⁸ These are supposed to represent more or less prototypical OMNs, diverging in e.g. number preferences or accessibility of their constituting elements. In addition, I analyse *habillement* ‘clothing’ which had to be excluded from the acceptability judgement study because of its outdatedness. However, since it seems to represent a rather prototypical SOMN given its deverbal origins and countability preferences (cf. Table 4.2), a diachronic analysis would be illuminating. Finally, the two collective nouns *équipement* ‘equipment’ and *accoutrement* ‘attire’ will also be examined. These two instances theoretically follow the same derivational pattern as *habillement* in originating as deverbal nouns on *-ment* and in oscillating between an eventive and an instrumental reading, but they do not represent OMNs. Semantically, they come under the same domain of artefactual functional nouns denoting things for carrying out a particular action, just like *habillement*. Syntactically, they are easily combinable with the indefinite article and show a clear possibility of occurring in the plural. Fr. *équipement* appears in 49% of the occurrences in *FrTenTen12* in the plural, and 5% of occurrences represent combinations with the indefinite article. Fr. *accoutrement* occurs in 23% of the occurrences in *FrTenTen* in the plural and in 7% of occurrences when it is combined with the indefinite article. It would therefore be illuminating to identify what kind of changes they underwent and at what kind of developmental step they stopped.

For each of these nouns I searched in *Frantext* for the respective lemma in every century from the first occurrence. I searched for the medieval spelling (*lexique médiéval*) up until the 16th century. Since it is not practicable to analyse the total number of occurrences for every noun under study, I created for each century and lemma a random sample from the statistically significant sample size (confidence level: 95%, margin of error: 5%). For example, there are 8,512 occurrences of *gens* ‘people’ in

⁶⁸ The collection nouns *mercerie* and *volaille* were not examined in the systematic main part of the acceptability judgement study, but were analysed in the exploratory part which focused on the distributive predicate ‘one after the other’.

the 15th century; the random sample then comprises 368 occurrences. In the case of those centuries where the overall number of occurrences was relatively low, I analysed at least 150 cases. This number is to a certain extent random but represents a manageable quantity to analyse. For instance, *accoutrement* ‘attire’ is found in 184 cases in the 20th century in *Frantext*, the statistically significant sample size is 125. In cases like these I thus increased the number of occurrences to 150. Furthermore, in many cases, the absolute quantity per century was under 150, e.g. *habillement* is only found in 27 occurrences in the 21st century. Here I analysed all existing cases and did not create a random sample. Table 8.1 below shows the absolute and relative frequencies (extrapolated to 1 million) of all nouns under study for every century in the *Frantext* corpus, as well as the calculated random sample sizes (ss). Some of the occurrences in the samples did not represent relevant cases for my analysis. Particularly in the earlier centuries, the noun tagging did not work accurately so that the sample also included cases of e.g. the adverb *habillement* ‘skilfully’ or the Old French adjective *gent* ‘noble’. Those cases were excluded from the analysis. However, I considered every nominal occurrence of these nouns, even if it did not have a referential interpretation, but e.g. an eventive one, to capture possible meaning changes as assumed. Another reason for exclusion were double occurrences. In some few cases two editions of one and the same original document were found in the sample, so any repetitions were excluded. Finally, I excluded every non-referential occurrence (e.g. idiomatic/fixed expressions, metalinguistic comments), since these do not necessarily reflect the same kind of syntactic and semantic use as the referential cases. I did not exchange those cases given the presumption that the proportion of irrelevant cases in the sample also represents the same proportion in the whole population of occurrences. In Table 8.1, the sample sizes (ss) indicated represent the corrected quantities without the irrelevant occurrences. The table displays the absolute frequencies as well as the relative frequencies per million.

Parallel to the examination of the characteristics of collection nouns in present-day language use, I will also focus on the two aspectual criteria of external boundedness and internal plurality in language evolution. As was elaborated in chap. 3.1 and tested in chap. 4, the feature of external boundedness is reflected by nominal number and determination in Romance languages, with the feature of internal plurality characterised by various phenomena like *constructions ad sensum*, the combination with stubbornly distributive predicates and different hierarchical relations between the collection and its constituting members. In the following, I will shortly summarise these results and findings of the synchronic examination with focus on French to deduce from them the criteria for the present diachronic analysis.

External boundedness: Different kinds of collection noun may occur in the singular and the plural (collective nouns), only in the singular with no equiv-

Table 8.1: Frequencies of the collection nouns in *Frantext* and sample sizes analysed in the diachronic study.

	century	12th	13th	14th	15th	16th	17th	18th	19th	20th	21st
inanimate	noun										
	<i>fringues</i>	/	/	/	/	/	/	/	/	/	295 (2.8) ss: 166
	<i>mercerie</i>	/	/	15 (3.8)	16 (3.8)	28 (3.9)	12 (0.6)	7 (0.2)	85 (1.3) ss: 83	262 (2.4) ss: 154	26 (1.9) ss: 23
	<i>mobilier</i>	/	/	/	/	/	/	31 (1.0)	815 (12.8) ss: 261	767 (7.2) ss: 255	106 (7.9) ss: 81
	<i>vaisselle</i>	18 (16.5)	/	71 (18.1)	85 (20.3)	47 (6.6)	111 (5.2)	111 (3.6)	446 (7.0) ss: 204	1,542 (14.4) ss: 304	245 (18.2) ss: 136
	<i>habillement</i>	/	/	5 (1.3)	158 (37.8)	195 (27.4)	232 (10.8)	496 (15.9)	312 (4.9) ss: 172	331 (3.1) ss: 174	27 (2.0) ss: 17
	<i>vêtement</i>	28 (25.6)	13 (9.7) ss: 12	106 (27.1)	102 (24.4)	160 (22.5)	248 (11.5)	740 (23.7)	3,248 (51.0) ss: 253	6,943 (64.7) ss: 342	1,258 (93.2) ss: 265
	<i>équipement</i>	/	/	/	/	/	/	13 (0.4)	81 (1.3)	2,289 (21.3) ss: 327	111 (8.2) ss: 102
	<i>accoutrement</i>	/	/	/	9 (2.2)	140 (19.7)	29 (1.4)	22 (0.7)	105 (1.7)	184 (1.7) ss: 150	24 (1.8) 21
animate	<i>volaille</i>	/	/	12 (3.1)	10 (2.4)	64 (9.0)	77 (3.6)	77 (2.5)	337 (5.3) ss: 179	595 (5.6) ss: 228	98 (7.3) ss: 84
	<i>bétail</i>	/	/	46 (11.8)	21 (5.0)	325 (45.7)	550 (25.6)	111 (3.6)	321 (5.0) ss: 109	1,131 (10.5) ss: 175	93 (6.9) ss: 80
human	<i>gens</i>	2,079 (1,902.0)	1,576 (1,174.3)	4,827 (1,233.2)	8,512 (2,033.6)	4,914 (690.3)	8,889 (413.5)	13,727 (440.0)	25,773 (404.5)	37,450 (349.2)	6,043 (447.6)
		ss: 301	ss: 281	ss: 328	ss: 359	ss: 349	ss: 368	ss: 374	ss: 379	ss: 380	ss: 283

alent plural form (SOMNs) or only in the plural with no equivalent singular form (POMNs). Correspondingly, collective nouns are countable, SOMNs are not countable and POMNs are countable only in a restricted manner. The external (un)boundedness of collections is also reflected in the (in)compatibility of the nouns representing them with certain determiners. Countable collective nouns may thus be combined with all determiners implying such countability like the indefinite article, numerals and other distributive determiners like Fr. *plusieurs* ‘various’, but not the partitive article. In contrast, SOMNs are not combinable with the indefinite article, nor with other distributive determiners, but they are combinable with the partitive article. POMNs like Fr. *fringues* ‘clothes’ or *gens* ‘people’ are not combinable with determiners implying boundedness and singularity like the indefinite article, but they are with those implying boundedness and plurality like *plusieurs* ‘various’. In an idealised model of the evolution of collection nouns (CCN > SOMN > POMN), one may thus assume an increase and then a predominance of the singular (CCN > SOMN) and then an increase and a predominance of the plural (SOMN > POMN). This development may be accompanied by a decrease in combinations with the indefinite article and distributive determiners implying plural and then a re-appearance of the latter. As has been shown in part II of this work, such an idealised model is rather improbable due to the fact that other influencing factors like analogy or the form and size of referents may interfere. Nevertheless, I assume these tendencies to reflect basically the diachronic development of collection nouns.

Before turning to the issue of the internal plurality of collections, a note on determiners is necessary. The coding of the mass-count distinction via different determiners in Romance languages has already been summarised in Table 1.2, but the diachronic corpus analysis on French does, however, necessitate a more fine-grained overview on this topic. Not only were some of them grammaticalised in later periods of language evolution (the partitive article e.g. develops its feature of marking mass syntax only in Middle French) but a simple binary distinction between mass and count determiners does not often seem to suffice either. This latter topic was already touched upon in chap. 1.2.1, where the bounding potential of the definite article, possessive pronouns and demonstrative pronouns was illustrated with examples like Germ. *das Wasser* ‘the water’. Are these determiners thus to be categorised as belonging to the count domain? To a certain extent, yes, since they contextually bound referents, but they do not necessarily change the aspectual type of the noun in question. Example (72) is illustrative here:

- (72) J’ai seize ans, j’apprends à rentrer les foin^s et à m’occuper du *bétail*, à nettoyer les auges et à soigner les bêtes. (*Frantext*: C. Arnaud, *Qu’as-tu fait de tes frères?*, 2010, p. 130)

‘I am sixteen years old, I learn to bring in the hay and to care for the *cattle*, to clean the troughs and to take care of the animals.’

Although the definite article bounds the referents of *bétail*, it does not transform it into a typical collective noun like *troupeau* ‘herd’. The cattle in the context still remains an unidentified mass of animals, of which we do not know the exact type, and which have no internal structure or hierarchy (one cannot e.g. imagine a leader of the animals in this context). Although the statement that the definite article, possessive and demonstrative pronouns are not restricted to the mass-count distinction may therefore be true, one also has to consider coercion effects often triggered by them. Since they do not change the aspectual type of the noun in question, I treat them as *neutral determiners*. The same holds for the complex determiner <*tout* + def. art./poss. pron./demonstr. pron.> ‘all the/his/this’. When combined with a mass noun, the quantifier *tout* reinforces the external boundaries already established by the determiners: “Les N massifs ne sont nullement interdits si l’exigence de limites se trouve satisfaite” (‘mass nouns are not at all prohibited [in these contexts] if the requirement of outer boundaries is met’) (Kleiber 1998, 91). Kleiber cites the example of *J’ai bu tout le vin* ‘I drank all the wine’ which automatically implies some kind of outer delimitation, probably the whole bottle of wine. The same holds for the plural form *tous les/ses/ces*. Without the quantifier, the plural itself may either trigger a distributive or a collective reading, irrespective of the kind of determiner. The exact interpretation is then effected by the predicate, e.g. *cueillir* ‘to pick (single objects)’ vs. *rassembler* ‘to gather’, and not the determiner. *Tous* does not change much (or even anything) in this, as Kleiber (2012, 230) points out:

De même que *les* peut signifier ‘tous les’ sans avoir pour autant le sens intrinsèque de ‘tous les’, de même *tous les* peut s’interpréter de façon distributive sans être intrinsèquement un quantificateur distributive.

‘Just as *les* can mean ‘all the’ without having the intrinsic meaning of ‘all the’, *tous les* can be interpreted distributively without being intrinsically a distributive quantifier.’

The quantifier *tout*, without any determiner, also triggers a binding effect, but with mass nouns, this seems only to be restricted to sortal readings. For homogeneous mass nouns like *water*, Flaux (1999, 480) states that both *tout* and *chaque* trigger the universal sorter having the interpretation ‘each kind of water’. She does not mention heterogeneous mass nouns in this respect, but the rare examples of OMNs in combination with *tout* or *chaque* in *Frantext* indicate exactly this. In example (73) *tout bétail* may only be interpreted as any kind of cattle/livestock, not as a single animal:

- (73) Coupépé à la fois, cest orge, en herbe, séché et serré au grenier, comme l'autre foin, est aussi bonne viande pour *tout bétail* en hyver. (*Frantext*: O. de Serre, *Le Théâtre d'agriculture et mesnage des champs*: t. 1, 1603, p. 305)
 'Cut at the same time, this kind of barley, harvested still green, dried and huddled in the attic, like the other (kinds of) hay, is also good food for *any (kind of) cattle/livestock* in winter.'

Following Mihatsch (2006, 130), these kinds of sortal reading are only possible if there is a well-defined set of subcategories and thus a hyponymic relation between the collection itself and its constituting entities. One may thus assume that these kinds of occurrence are only possible when the collection in question is (primarily) defined by a hyponymic and not by a meronymic relation between the collection itself and its constituting entities.

Note, however, that especially in the very beginnings of documentation of the French vernacular, this categorisation may not necessarily apply. In texts of the 15th century, there are e.g. occurrences of *toute gent* as in (74), which do not have the interpretation of every single nation, but of all people, being expressed by *toute la* in Modern French:

- (74) [. . .] je demande se le peuple et *toute gente* se rebelloit contre un tel seigneur (*Frantext*: C. de Pisan, *Le livre de la paix*, 1412, p. 123)
 '[. . .] I ask if the people and *the whole people* would rebel against such a sovereign'

In the case of <*tout* + indef. art.>, the sortal reading of *tout* is not possible and it exclusively triggers an interpretation of individuation with (heterogeneous) mass nouns – it thus functions here as a universal packager. Since *tout* only refers to the whole referent, *tout un* reinforces the outer delimitations of individual count nouns as in *J'ai mangé toute une pomme* 'I ate a whole apple' and of collective nouns as in *J'ai acheté toute une tenue* 'I bought a whole outfit' (cf. Kleiber 1998, 94). With OMNs I thus assume either the non-existence of this complex determiner, or that the combination then triggers a collective interpretation. In the latter case, the complex determiner will then be categorised as a count determiner.

Another issue is related to identifying adjectives like *autre* 'other', *tel* 'such', *même* 'same', *dernier* 'last' or *certain* 'certain'. These are not exactly determiners and often occur together with the definite or indefinite article, but they may syntactically be put into the position of one – they are thus classified as such in constructions like *(un) tel/autre vélo* 'such a bike/another bike'. These kinds of construction have two possible interpretations with countable nouns (individuals and collections): they

either refer to another (sub-)category of the referent in question (another kind of bike) or they refer to another individual of the same category (another exemplar of this kind of bike). With mass nouns they, like *tout*, exclusively trigger sortal reading (cf. Flaux 1999, 480). Again, Flaux (1999) only mentions homogeneous mass nouns like Fr. *eau* ‘water’ to illustrate her explanations, but since both heterogeneous and homogeneous mass nouns are uncountable, I assume the same tendencies for both noun categories. Those occurrences, where a noun is combined with one of these identifying adjectives are thus treated as follows: if they occur together with the indefinite article (*un autre N*), they will be classified as count determiners. If they occur with the definite article (*l’autre N*) or alone (*autre N*) they are first categorised as a separate group of determiners to be afterwards classified more specifically. The same holds for occurrences in the plural.

Other quantifiers and determiners are less ambiguous. All numerals as well as all distributive determiners like *quelques* ‘some’ and *plusieurs* ‘various’ are regarded as belonging to the count domain. This does not hold for the singular form of *quelque* ‘some’ which like *pas de* ‘no’, *tant de* ‘so much/so many’, *beaucoup de* ‘much/many’ or *assez de* ‘enough of’ quantify referents of count as well as mass nouns. These are therefore classified as non-distributive quantifiers belonging to the *neutral domain*. Binominal numerals like *une douzaine de* ‘a dozen of’ will be first classified as numerals. Possible differences in this respect will be discussed if or when they occur (the acceptability judgement tests presented in chap. 4 indicated no statistically significant differences between numerals and attenuating binominal numerals in combination with an OMN). The most unambiguous are, finally, the indefinite article which clearly marks nouns as belonging to the count domain and the partitive article that marks nouns as belonging to the mass domain (cf. Nicolas 2002, 5–6; Stark 2007; Vermote/Lauwers 2016; cf. also chap. 1.2.2 and Table 1.2). These latter two indefinite determiners with their present functions also have to be considered, however, in their diachronic evolution. Zero determination in Old French was still very common and not the exception as it is in Modern French (cf. Table 1.2) and the two indefinite determiners are only grammaticalised over the period of Old and Middle French (cf. Marchello-Nizia 1999, 76; Carlier/Lamiroy 2018, 145). Whereas we can thus rely on the indefinite and partitive article to overtly mark count and mass syntax in Modern French, this is not necessarily the case for older stages of language development (cf. Carlier/Lamiroy 2018, 151). In what follows, I will thus summarise the evolution of these two indefinite articles in French focussing mainly on the following two questions: when they arose and when they became obligatory and with which kind of noun could they be combined (mass, count, abstract nouns, for instance). The other determiners and quantifiers discussed so far did not undergo crucial semantic or syntactic changes relevant for this present analysis (cf. Marchello-Nizia 1999, 71).

The indefinite article in French *un(e)* goes back to the Latin numeral *ŪNUS*. As such, it is still often used as a numeral in Old French and only from the 14th century onwards does it start to function as an article introducing indefinite referents first with specific and then also with non-specific and generic reference (cf. Carlier 2001). In all cases, however, it needs a bounded referent to combine with. The absence of the indefinite article in Old French texts thus does not necessarily indicate unboundedness, but in every stage of evolution its presence expresses boundedness. Carlier (2013, 49) finds that “*uns* is also compatible with non-count nouns, including mass nouns and abstract nouns”. She cites two examples which, however, in my view represent cases of sortal readings, i.e. *un marbre fin blanc et bis et si bel* ‘a fine, white and greyish marble’ as well as *une amour si tres pure* ‘such a pure love’. I therefore categorise contexts without the indefinite article as neutral to the mass-count distinction, but every instance of the indefinite article will be categorised as indicating count syntax. The latter contexts will then be subclassified as referentially individuated or as sortal readings. This procedure allows me to distinguish between occurrences where the noun indeed refers to a bounded entity, an individual or a bounded collection and those where the sortal reading does not necessarily indicate a different kind of aspectual type. Sometimes, such a distinction, however, cannot be clearly made. A preliminary example from my corpus analysis may illustrate this. In example (75), Fr. *équipement* is combined with an indefinite article. The prepositional phrase *de premier ordre* ‘excellent’ indicates a qualification of the referents and thus implies a sortal reading. One may then interpret the example either as a SOMN which is contextually bounded through qualification or as a CCN where the qualification does not change its constitution.

- (75) Il se montre fort pessimiste et j’ai passé près d’une heure à essayer de lui démontrer que l’armée américaine avait *un équipement de premier ordre* et que nous pouvions compter sur elle. (*Frantext*: J. Green: *Journal*, 1943, p. 265)

‘He is very pessimistic and I spent almost an hour trying to show him that the U.S. Army has (*an*) *excellent equipment* and that we can count on it.’

Interestingly, Old French also had a plural form of this indefinite article *uns* which disappeared in the 15th century. This determiner was not an exact one-to-one mapping from a single bounded entity to many single bounded entities – thus no distributive plural marker – but rather a determiner of a cohering plural like *uns mariniers* ‘some sailors’, who are tied together by some internal or external factor (cf. Herslund 2003; Carlier 2016). The determiner is, however, very marginal in my corpus analysis, so will take no further part in this discussion.

The French partitive article *du/de la* originates as a real partitive introducing portions of an aforementioned referent. Consequently, there is in Old French the syntactic and semantic distinction between *mangier pain* ‘to eat (some) bread’ and *mangier del pain* ‘to eat from the aforementioned specific quantity of bread’ (cf. Carlier 2004, 130) – the partitive *del* thus does not express an indefinite mass, but an actual portion. As such, it is typically found in contexts with verbs like *prendre* ‘to take’, *boire* ‘to drink’ or *manger* ‘to eat’ and in the beginnings only with concrete mass nouns (cf. Englebert 1996, 13–14; Carlier 2004, 127). From the 15th century on, there are occurrences with the partitive article in combination with abstract nouns like *ennui* ‘boredom’ (cf. Englebert 1996, 15), and from the 17th/18th century onwards these occurrences become common (cf. Carlier 2007, 28). Instances of the partitive article consequently are not always necessarily indicators of mass syntax. Especially at the very beginning of the evolution of this kind of determiner, there is the possibility of ‘real’ partitives determining a bounded portion like a loaf of bread or a bottle of wine. In the corpus analysis, partitives articles are thus not automatically categorised as mass determiners but will be interpreted one by one in context. Only from the 15th century onwards can we speak of a partitive article in contrast to the partitive prepositional construction in Old French (cf. Carlier 2007).

Just like the indefinite article, the partitive article also developed a plural form, in this case *des*. It displays a parallel evolution to its singular counterpart, but, for being morphologically plural, it first denotes a specific set of a plurality of entities and then becomes a marker of the indefinite plural (cf. Carlier/Lamiroy 2014).

The classification system described here is summarised in Table 8.2. Determiners triggering any kind of sortal reading are classified as being somewhere in between the count and the neutral domain. It is true that classes and categories also represent bounded sets of (possible) referents. These bounding effects, however, do not result in holistically perceptible entities, but only hypothetical referents sharing the same properties.

Table 8.2: Classification of determiners and quantifiers in the diachronic analysis.

neutral	count	mass
definite articles	indefinite articles (+ identifying	partitive
possessive pronouns	adjectives/+ <i>tout</i>)	articles
demonstrative pronouns	<i>plusieurs/quelques</i>	(from
non-distributive quantifiers (e.g. <i>beaucoup de</i>)	numerals	15th cent.
< <i>tout/tous</i> + def. art./poss. pron./demonstr. pron.>		onwards)
(definite article +) identifying adjectives		
<i>tout/chaque</i> (sortal readings only)		

Internal plurality: With respect to the internal plurality of collection nouns, chaps. 3.1 and 4 addressed the issues of the accessibility of the constituting entities of a collection as well as its constituting hierarchical relations. Accessibility of internal plurality may be reflected by a number of constructions. The acceptability judgement tests focused on the compatibility of different kinds of collection nouns with stubbornly distributive predicates and highly distributive constructions like ‘one after the other’. The results showed that accessibility of the internal plurality of a collection generally increases from collective nouns to OMNs, but also that different constructions are susceptible to this kind of combinability to varying degrees. In this respect, the highly distributive equivalents of ‘one after the other’ were more restrictive than the tested stubbornly distributive predicates. I will thus categorise every modifier of the noun in question. This allows me to identify each kind of distributive predicate and to classify consequences for interpretation in terms of the nominal aspectual type in question. Chap. 3.1.2 also addressed the issue of *constructions ad sensum*, possible with human collections in English, but highly restricted in French and in other Romance languages. In the framework of the synchronic analysis, I already indicated that there is no possible way to uniformly analyse these kinds of constructions, since *constructions ad sensum* are governed not only by the mere combination of noun and predicate, but also by the distance between the two elements or the diamesic variety of the source text (conceptually spoken vs. conceptually written). I will nevertheless categorise for every occurrence its syntactic position, and in the case of a subject, the number of the predicate. I expect the indication of the degree of accessibility of the internal plurality of a collection by means of these constructions not to be of high frequency, as the corpus data of present-day French has already indicated (cf. chap. 4.1). For this reason, I will consider the kinds of adjectives the collections are combined with, the number of the verb when the collection under study is in subject position, as well as other conspicuous constructions reflecting the internal plurality of the collections as part of a more qualitative rather than systematic-quantitative study.

The hierarchical relations between the constituting entities of a collection and the collection itself are linguistically indicated by several binominal constructions like *kind/type of X* or *part/member of X* as well as coordinating constructions like *X and other Ys* or various verbs like *X is constituted of Ys* (cf. chap. 3.1.2). The latter two constructional types served well in the acceptability judgement tests to show that most OMNs are indeed constituted by both a meronymic and a hyponymic relation between the collection and its constituting parts, but also that they are not prototypical candidates for holonyms or hyperonyms (cf. chap. 4.2.2). To analyse these hierarchical relations in the diachronic evolution of collection nouns, I recorded the verbs they occur with and whether these indicate

e.g. a meronymic relation as well as whether the noun in question occurs as an N2 in a binominal construction and whether the N1 then indicates the semantic type of relation (e.g. *type de* ‘type of’). In addition, I will also report contextual aspects which provide clues to the composition of the constituting entities, whether they are more homogeneous or rather heterogeneous, i.e. whether *mobilier* ‘furniture’ e.g. refers only to wooden furniture or to possessions in general. This could also be an indicator of either a hyponymic relation between the collection and its constituting entities leading to the collection being governed by similarity, or possibly a meronymic relation for its being constituted by contiguity. This aspect has to be considered, however, only as an additional, indirect piece of evidence since e.g. heterogeneous reference does not always and automatically imply spatio-temporal contiguity, but could also indicate functional similarity.

Summarising, I will in the following analyse the occurrences of collection nouns as displayed in Table 8.1 from their recorded beginnings to present-day French. In a first step, I concentrate on the etymology and possible ways of coming into being of a collection noun as assumed in Figure 7.1. With this, I reconstruct not only the possible origins of collection nouns as described by Baldinger (1950), but I will also address the assumption of i.a. Grimm/Levin (2017) that at least prototypical OMNs are artefact nouns tied to a certain kind of event (chap. 8.2). In a second step, I trace the further evolution of the collection nouns under have taken in the *Frantext* corpus, categorising them by means of possible pathways they took (chap. 8.3). The results of these two steps taken together will then serve to address and discuss the research questions formulated above (chap. 9).

8.2 The etymology of collection nouns

The state of the art on the coming into being of collection nouns was summarised in Figure 7.1 as three main pathways of evolution. Following this, collection nouns originate either as entities involved in some kind of event or action, as entities being constituted by one common property or as the fossilised Latin neuter plural. The collection nouns focused on in this analysis can be classified as exemplars for all three types, although the categorisation may not always be as obvious as seems at first sight.

The French derivates on *-ment* are to be classified as prototypical examples of the first lexicalisation path. Fr. *habillement* (< MFr. *habiller* ‘to equip’; cf. FEW, s.v. **bilīa*), *vêtement* (< Lat. *vestimentum* < *vestīre* ‘to dress’; cf. FEW, s.v. *vestimentum*), *équipement* (< OFr. *eschipier* ‘to equip a ship with crew and gear’; cf. FEW, s.v. *skipa*) and *accoutrement* (MFr. *acoustrer* ‘prepare, adorn’; cf. FEW, s.v. **consūtūra*) are all deverbal derivates denoting in their origins the instrument of a

certain action. In the terminology of Grimm/Levin (2017) they are functional artefact nouns denoting objects made for a potential action (in contrast to stage-level artefacts that have already participated in a certain action, cf. e.g. Engl. *mail*). In this sense, *habillement* originally denotes ‘what serves to dress sb.’, *vêtement* ‘what serves to cover sb.’s body’, *équipement* ‘everything that serves to equip sb. for some task’ and *accoutrement* meant ‘what serves to dress/attire sb.’. All these actions of equipping, dressing and attiring normally imply a plurality of objects so that foundations for the development of a collection noun are laid. This implication of more than one object being involved in that particular action is, however, not necessarily present, as the example of Lat. VESTĪMENTUM/Fr. *vêtement* shows. The Latin etymon and the French continuation both may refer to a single garment. Other collection nouns do not show this very clear relation to an action in being deverbal derivatives formed by means of a particular suffix, but they nevertheless display some binding to an event. This is the case with Fr. *mercerie* ‘haberdashery’ which is built on the base of *mercier* ‘haberdasher’ and thus denotes the items sold by a haberdasher (cf. FEW, s.v. *merx*). In this sense, *mercerie* is not primarily a functional artefact noun (‘things that serve for sewing and other manual work’), but a stage-level artefact noun (‘things sold by a haberdasher’). *Fringues* is a deverbal noun going back to *fringuer* ‘to prance’ > ‘to fit out’, which in turn supposedly originates in the onomatopoeic *fring-* ‘hop’ (cf. FEW, s.v. *fring-*). Although *fringues* thus does not show the same derivational pattern as e.g. *accoutrement*, the meaning of a functional artefact noun is similar in equally denoting the objects serving to fit somebody out.

The two animate collection nouns *bétail* ‘cattle’ and *volaille* ‘poultry’ are cases of the third lexicalisation path – the fossilisation of the Latin neuter plural – although the path is not direct. French nouns on *-aille* denoting animals mostly represent cases of analogy to OFr. *aumaille* ‘cattle/head of cattle’ (cf. Baldinger 1950, 130). This collection noun goes back to the inflectional plural of Lat. ANIMAL, which is *animalia*, and survives in nearly all Romance languages, although not always as a collection noun (e.g. Sp. *alimaña* ‘animal’, Rom. *nămaie* ‘horned cattle’) (cf. FEW, s.v. *animal*). Fr. *volaille* thus goes back to Lat. VOLĀTILIA, neuter plural of *volātīlis* ‘winged/feathered’. Regular sound change led to *volille* in Old French, but through analogy it then changed into *volaille* (cf. FEW, s.v. *volatilis*). A similar evolution can be seen in *bétail*, which goes back to Lat. BĒSTIA ‘animal’. The Latin etymon and also its regular continuation OFr. *beste*/ModFr. *bête* actually have no plural meaning; the meaning of a collective plural was only possible through suffixation by means of *-aille* on the model of *aumaille*. In contrast to *volaille*, *bétail* has not the feminine but the masculine gender. Until the 14th century, there is thus *bestiaille* (fem.) and since the 13th century *bestial* (masc.) (cf. Baldinger 1950, 130). Both animate collection nouns designate a collection of

certain animals, linguistically cohered by the suffixation with *-aille*. Fr. *vaisselle* represents a case of regular sound change from Lat. *VASCELLA* (neut. pl.) ‘small recipients’. It is thus a clear case of the direct fossilisation of the Latin neuter plural, without any detours like in the case of *bétail* and *volaille*. Note that the inflectional singular Lat. *vascellum* regularly led to a singular object noun, *vaisseau* ‘recipient/ship’ (cf. FEW, s.v. *vascellum*).

The two remaining collection nouns *mobilier* and *gens* do not completely fit the picture of the assumed lexicalisation paths. Fr. *mobilier* ‘furniture’ goes back to the adjective *mobilier* (< *mobilier* ‘what concerns the movable property’ < *mobile* ‘movable’ < Lat. *mobilis* ‘movable’) (cf. FEW, s.v. *mobilis*). It thus very likely represents a case of lexical absorption where in constructions like *biens mobiliers*, *effets mobiliers* ‘movable property’ the meaning of the whole construction is transferred to the then nominalised adjective. One may thus assume some kind of property-orientated lexicalisation path (path II.), which is, however, not as prototypical as Germ. *Mannschaft* ‘state of being a man/relation between the overlord and his vassal’ > ‘team’ for being involved in additional derivational processes like lexical absorption and conversion. The same may hold – very broadly speaking, for Fr. *gens*. This collection noun goes back to Lat. *GĒNS* meaning ‘tribe, lineage’; there is thus a collection noun already in the Latin origin. The members of this collection share the property of being of the same origin. This original meaning is still preserved in Old and Middle French, often with some pejorative connotation – e.g. the barbarian tribe in contrast to the Christian people. In contrast to other Romance languages, French has developed a plural form having first the meaning of people standing in relation to some central person (liegemen, servitors) to then becoming the denotation of people as a POMN (cf. FEW, s.v. *gens*).

The nouns investigated here thus fit the assumed pathways of lexicalisation of collection nouns well, although the path may not always be as direct as the models suggest. In what follows, I will trace the diachronic development of all the collection nouns under study classifying them on the basis of the three lexicalisation pathways discussed in this section. Because countability preferences, those determiners the noun is combined with as well as all relevant semantic properties marked by adjectives, verbs or potential referents named in the context are mutually dependent in all reflecting together the aspectual type of the noun in question, a structuring of the analysis in line with the features of external boundedness and internal plurality (as was done in chaps. 3.1 and 4) would not be very fruitful in this diachronic section. So I will first address the event-related exemplars of the first pathway, then the property-related exemplars of the second pathway to finally come to those nouns related to the Latin neuter plural, before finally discussing all features analysed together.

8.3 Results of the corpus analysis

8.3.1 Event-related artefact collection nouns: *habillement*, *accoutrement*, *vêtement*, *équipement*, *mercerie*, *fringues*

The preliminary analyses regarding the synchrony of present-day language revealed that *habillement* nowadays is a typical SOMN (it occurs only in the singular and it is not compatible with the indefinite article, cf. Table 4.2), *fringues* is a POMN (it occurs mostly in the plural and is not compatible with the indefinite article, but to some extent with other distributive quantifiers implying plurality, cf. chaps. 4.2.1 and 4.2.2), *accoutrement* and *équipement* are CCNs (they denote bounded collections, may occur in the plural and are easily combinable with the indefinite article, cf. *supra*), *vêtement* is a flexible noun that is used mostly to denote individual objects, but can be used as a SOMN (cf. chap. 4.2.2) and *mercerie* may be used as a SOMN as well, but it also denotes a haberdashery shop (cf. Table 4.2). For all these collection nouns I assume different syntactic and semantic evolutions according to their current aspectual type and depending on the assumed lexicalisation path of collection nouns that was summarised in Figure 7.2. The SOMNs *habillement* and *mercerie* should display an increasing number of uses in the singular which should at some point become the predominant usage. This increase in the inflectional singular should accompany a decrease in combinations with distributive quantifiers as well as a stabilisation of the semantic features and a homogenisation of the constituting referents. The POMN *fringues* should, following the assumed path of lexicalisation of collection nouns, display a similar evolution but going one step further than the SOMN. This development should manifest itself by an increase in and ensuing predominance of the inflectional plural, and any kind of distributive quantifier should be limited to the plural. The evolution of the two CCNs *accoutrement* and *équipement* should stop at some point before the phase when their referents have become homogenised and the collection noun thereby syntactically massified. It may also be that the very first steps along the path are documented for these two collection nouns, i.e. probably with eventive meanings and serving as functional nouns. It is not exactly clear what to expect for Fr. *vêtement*. The fact that it is a direct continuation of the Latin etymon leads to the assumption that *vêtement* may already have undergone the main steps of evolution in Latin or before the first attestations in the French language as documented in the corpus. I thus assume for *vêtement* a post-SOMN stage in Old French and its further development as suggested by Mihatsch (2016), since the SOMN stage of ‘clothing’ has already been attested for Lat. *vestimentum*. Unfortunately, there is little information about Lat. *vestimentum* and its origins. Etymological dictionaries of Latin mostly only comment on the Indo-Germanic origins of the root *vest-*, but nothing further on this specific noun.

This is unfortunate in more than one way since Lat. *vestis* is indicated as ‘clothing’, but nothing more is given (cf. LEW, s.v. *vestis*).⁶⁹ In what follows I will analyse these assumptions for each functional artefact collection noun under study.⁷⁰

Habillement: Fr. *habillement* ‘clothing’ is first attested in the *Frantext* corpus in the first half of the 14th century. As shown in Figure 8.1 below, these initial attestations are predominantly in the plural, but over the centuries there has been a constant increase in singular forms until the 21st century when the inflectional plural is no longer attested in the corpus. Since the analysis takes into consideration only manageable sample sizes for each century, a simple comparison of their relative frequencies – which also considers the evolution in overall frequency – is unfortunately not possible. Instead, the following presents the quantities of singular vs. plural, mass vs. count determiners etc. proportionally. The relative frequencies are given in the respective data tables under the graph to give the full picture. The analysis of the proportions between the inflectional singular and plural as displayed in Figure 8.1 confirms the first hypothesis of an increase in occurrences of the inflectional singular. An examination of the relative frequencies as displayed in the data table additionally indicates a decrease in the overall frequency of the collection noun.

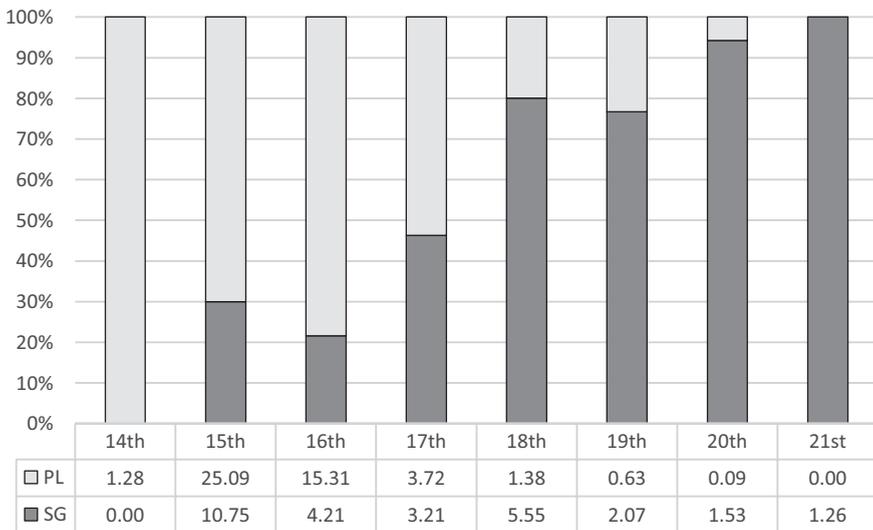


Figure 8.1: Evolution of number inflection of Fr. *habillement* in *Frantext* (rel. frequencies extrapolated to 1,000,000).

⁶⁹ Unfortunately, the extensive *Thesaurus Linguae Latinae* (TLL) has not yet reached the letter V.

⁷⁰ See <https://hdl.handle.net/11022/0000-0007-EB8E-1> for all results.

The plural occurrences of *habillement*, predominating until the 17th/18th century, are uniformly used to denote individual objects, either items of (war) equipment (cf. (76a)) or pieces of clothing (cf. (76b)) – the former meaning, however, ceases to be documented in the second half of the 16th century. The other meaning of ‘garments’ survives with decreasing frequency and continues into the 20th century (cf. (76c)).

- (76) a. Et pour la haste que eusmes, ne veymes point une salle playnne de blans harnois de guerre comme cuiraches et autres *habillemens*, et bringandines couvertes de drap de soye. (*Frantext*: G. Lengherand, *Voyage de Georges Lengherand*, 1486, p. 82)
 ‘And because we were in a hurry, we did not see a hall full of white armour like cuirasses and other *items of (war) equipment*, brigandines covered with silken cloth.’
- b. Le marchand, le voyant bien habillé d’*habillemens* precieux, dist [. . .] (*Frantext*: Anonymous, *Le violier des histoires rommaines moralisées*, 1521, p. 146)
 ‘The merchant, seeing him well dressed in precious *clothes*, said [. . .]’
- c. Ils étaient tous voués, alors, et jusqu’à deux ans leurs *habillements* n’avaient d’autres couleurs que le blanc et le bleu. (*Frantext*: H. Pourrat, *Les Vaillances, farces et aventures de Gaspard des montagnes*, t. 2, 1925, p. 183)
 ‘They were all damned, then, and until the age of two their *clothes* had no other colors than white and blue.’

Except for this semantic stabilisation on items of clothing, there is not much change in the use of the plural forms of *habillement*. It is mostly used with non-distributive quantifiers and either without any modifier or with mostly qualitative adjectives like *royal* ‘royal’, *beau* ‘beautiful’ or *riche* ‘rich’. More importantly, there are crucial changes in the semantics, the syntactic compatibilities and the contextual uses of the inflectional singular. In its beginnings, *habillement* had, according to its etymology and parallel to the inflectional plural, quite a broad range of possible meanings. The original meaning, tied to the base verb *habiller* is seen most clearly in (77a) where there is no relation to clothing at all, but only to a very general understanding of equipment. Most occurrences found in the corpus range over individual objects serving to equip or dress somebody (cf. (77b)), the bounded equipment or attire of somebody (cf. (77c)) and highly marginally also as clothing in general. Possible SOMN contexts as in (77d) may either refer to the complete outer appearance of the person in question or only to his/her clothing. An eventive meaning is not attested in these first centuries.

- (77) a. Vous avés tout l'*habillement* A taconner le garnement: Plume, papier, encre, marteaulx Enclume, tabule, marteaulx! (*Frantext*: Anonymous, *Le Sermon des barbes et des brayes*, 1450, p. 257)
 'You have all the *equipment* to mend the gear: Feather, paper, ink, hammer and anvil, slate, hammer!'
- b. [. . .] sur lequel cheval, bien chevauchant, il estoit armé de toutes pieces, reservé son *habillement de teste*. (*Frantext*: A. de la Vigne, *Le voyage de Naples*, 1495, p. 304)
 '[. . .] on this horse, riding, he was armed with all the pieces (of his armour), except for his *helmet* (lit. item of head equipment).'
- c. Et ce jeune duc print ung *habillement* de François, et partit, luy deuxiesme seulement, pour se retirer en son pays. (*Frantext*: P. de Comynnes, *Mémoires*, t. 2, 1489, p. 3)
 'And this young duke took a *costume* of a Frenchman and left, him only second, to retire in his country.'
- d. Ils le recogneurent à son *habillement* et langage, [. . .] (*Frantext*: B. de Vignère, *L'histoire de la décadence de l'Empire grec, et établissement de celui des Turcs*, 1577, p. 490)
 'They recognised him by his *outer appearance/clothing* and his language, [. . .]'

Just as with occurrences of the plural, over time there is a semantic specification towards 'things to dress sb.' and the original meanings of 'equipment', 'armour' and 'tools' are only marginally documented from the 16th/17th century on. As shown by (77c) and (77d), *habillement* originally comprised everything that serves outer appearance, i.e. clothing, but also accessories, jewellery, hair style, shoes and so on. After being semantically homogenised to clothing, there is from the 18th century an additional specification of the referents to outer clothing excluding additional items. As a consequence, *habillement* occurs in the corpus in conjunction with referents other than clothing:

- (78) a. Sa coëffure et son *habillement* répondoient à la simplicité de tout son extérieur, [. . .] (*Frantext*: C. Godard d'Aucour, *Thémidore*, t. 2, 1744, p. 19)
 'Her hairstyle and her *clothing* corresponded to the simplicity of all of her outer appearance, [. . .]'
- b. L'administration avait coutume de passer un marché pour la chaussure et l'*habillement*; [. . .] (*Frantext*: H. de Balzac, *Louis Lambert*, 1846, p. 609)
 'The administration used to make a contract for shoes and *clothing*; [. . .]'

In Modern French, *habillement* is now mostly restricted to lexicalised binominal constructions like *magasin d’habillement* ‘stockroom in military bases, prisons, concentration camps’ or *capitaine d’habillement* ‘quartermaster, i.e. person who supervises the (supply of) clothing, uniforms, equipment in the army’ and predominantly occurs in texts relating to the world wars. Parallel to this feature of semantic homogenisation and the increasing appearances in the inflectional singular, the singular forms are decreasingly combined with distributive quantifiers implying count syntax from the 17th century on. In the 21st century, there are no longer any occurrences of *habillement* with a distributive quantifier:

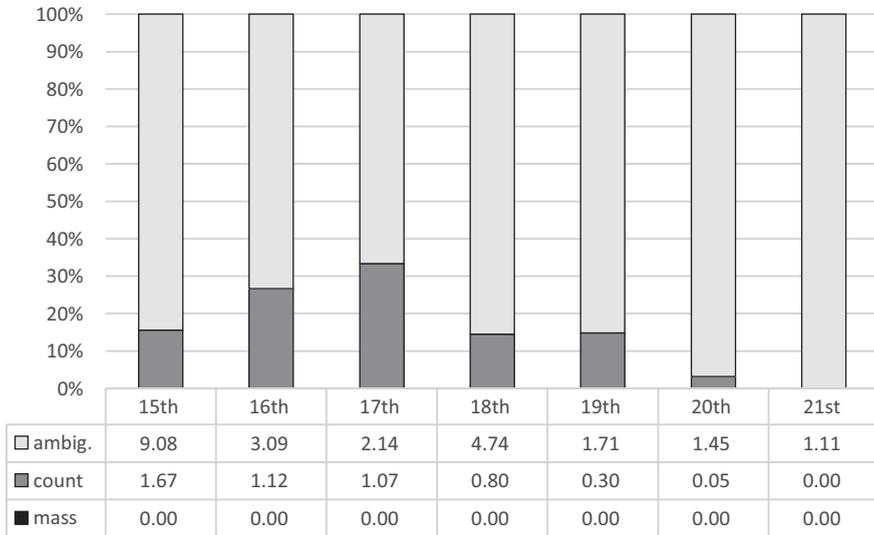


Figure 8.2: Evolution of types of determiner in combination with Fr. *habillement*_{sg} in Frantext (rel. frequencies extrapolated to 1,000,000).

Most of the occurrences of *habillement* with a count determiner are instances with the indefinite article. These are then mostly contexts where only an interpretation of *habillement* as ‘attire’ (cf. also (77c) and (78a)) is possible, but also examples of *habillement* as ‘garment’. Although, as shown in Figure 8.2, there are so far no occurrences of *habillement* in combination with a partitive article, but the 19th century sees the first attestations of classifier constructions, clearly indicating its use as a syntactic mass noun:

- (79) [. . .] puis, un professeur brutal, affublé de sa robe noire et de son bonnet carré, *pièce d'habillement* que je n'ai jamais pu voir sans humeur. (*Frantext*: E.-J. Delécluze, *Journal: 1824–1828*, 1828, p. 210)
 '[. . .] and then, a brutal professor, dressed in his black robe and with his square hat, *piece of garb* that I never could see without humour.'

In the case of *habillement*, there is a constant increase in inflectional singular forms, first an increase and then a decrease in distributive quantifiers as well as a semantic homogenisation of both the singular and the plural forms. Most occurrences in the corpus are, however, not clearly determinable by means of a mass or count determiner or classifier constructions like in (79). To finally determine the actual nominal aspectual type of each occurrence of *habillement*, the context consequently plays an important role: this comprises e.g. co-occurrences and cases of conjunction with unambiguous cases of other collection nouns, heterogeneous or rather homogeneous referents named as exemplars (cf. e.g. (77a)), and binominal constructions like the ones mentioned above, where *habillement* without a determiner implies (quasi-) generic reference that suggests the use as a SOMN and so on. Many occurrences are, however, not unambiguously determinable and are thus coded as *ambiguous*. This applies not only for this present analysis of *habillement*, but, as will shortly be shown, also for the other nouns examined. Figure 8.3 portrays the proportional evolution of these aspectual types of *habillement*:

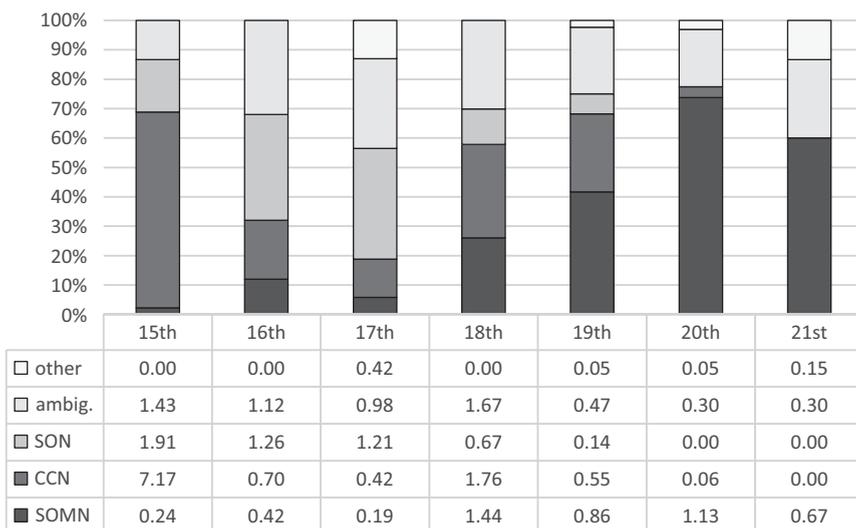


Figure 8.3: Evolution of nominal aspectual types of Fr. *habillement*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000).

As shown in Figure 8.3, Fr. *habillement* represents a nearly perfect example of the assumed pathway of lexicalisation of collection nouns: it originated as a deverbal noun denoting a functional category of things that serve to equip or dress somebody or something. The noun then becomes a CCN denoting somebody's armour or attire to finally become a SOMN which is not countable anymore and which is then restricted to a person's outer clothing. As shown in Figure 8.3, however, there is a relatively consistent proportion of ambiguous cases, which may also be interpreted as either singular object nouns (SONs) (cf. (80a)) or as CCNs (cf. (80b)). In addition, there are in nearly every century very marginal cases of *habillement* which may be interpreted as an action noun denoting the event of dressing up (cf. (80c)).

- (80) a. Les ecclésiastiques, de quelque nation qu'ils fussent sortis, dûrent aussi conserver toujours l'habit long, ou la toga, parce qu'il étoit *l'habillement* d'un citoyen romain. (*Frantext*: abbé J.-B. Dubos, *Histoire critique de l'établissement de la monarchie françoise dans les Gaules*, t. 1, 1734, p. 604)
 'Religious people, no matter what nation they came from, should also always keep the long robe, or toga, because this was the *garment* of a Roman citizen.'
- b. Ce même jour, dans l'après-midi, nous reçûmes *l'habillement* du bataillon. (*Frantext*: É. Erckmann, *Le conscrit de 1813*, 1864, p. 80)
 'The same day, in the afternoon, we got the *equipment/uniforms_{coll}* for the battalion.'
- c. Leur *habillement* progresse lentement parce que nous sommes dans des difficultés incroyables à ce sujet. (*Frantext*: C. de Gaulle, *Mémoires de guerre*, t. 3, 1959, p. 343)
 'Their *fitting out* advances slowly because we have a lot of difficulties with respect to this.'

Finally, the fact that there are plural occurrences attested from their beginnings as well as the fact that these all refer to individual entities and not to sets or collections leads to the assumption that there is a parallel evolution of *habillement_{SG}* and *habillements_{PL}*. As assumed in Figure 7.1 and 7.2, artefactual collection nouns originate as action nouns. The consideration of the etymology of various collection nouns in the framework of this present analysis, but also in chap. 3.2, leads to the conclusion that the intermediate stage between these two has to be very broadly assumed as 'objects involved in an action'. In the case of artefactual nouns, the kind of involvement is mostly either the result of that action or is an instrument for effecting this action as in the case of *habillement*. At least regarding these instrumental nouns, the plural occurrences are thus the pre-stage of a collective noun: provided by the right semantic frame, in this case preparations

for war, the loose plurality may be bounded to become a set of objects. As shown in Figure 8.1 these two aspectual types may very well coexist for a long time.

Mercerie: A similar evolution should in theory be undergone by Fr. *mercerie* ‘haberdashery’ – this is, however, not the case. Ever since its very first attestations in the 14th century, *mercerie* in the singular has oscillated between a locative sense such as a shop where small goods are sold (cf. (81a)) and the small goods themselves (cf. (81b)). At the same time, from the very beginning there are also plural occurrences designating small goods taken individually and not as an indefinite mass (cf. (81c)). In all these three interpretations, *mercerie* is in its meaning still close to the Latin etymon *MERX* ‘good/product’. Only from the 17th/18th century on are there occurrences of *mercerie* in the present sense of haberdashery (‘shop’ and ‘sewing wares’) (cf. (81d)).

- (81) a. et vit, lui qui parle, que ledit Joesne acheta en la *mercerie* du Palais deux anneaux d’argent dorez [. . .] (*Frantext*: Anonymous, *Registre criminel du Châtelet*, t. 2, 1389, p. 44)
 ‘and he who speaks saw that the aforementioned Joesne bought in the *small goods shop* of the palace two gilded silver rings [. . .]’
- b. [. . .] il seul a prins et gaigné, d’une vielle merciere qui vent *mercerie* derriere Saint-Innocent, XII coiffes de soye, [. . .] (*Frantext*: Anonymous, *Registre criminel du Châtelet*, t. 1, 1389, p. 66)
 ‘[. . .] he only took of an old haberdasher, who sells *haberdashery* behind the cemetery of Saint Innocent, 12 silken bonnets, [. . .]’
- c. Et nous vinrent veir [. . .] Sarasins qui nous apportoient *merceries* et toutez cosses a vendre; [. . .] (*Frantext*: C. de Velaines, *Le pèlerinage du tounaisien Coppart de Velaines en Terre sainte*, 1432, fol. 17v^o)
 ‘And there came Saracens to see us, who brought *small goods* and various things to be sold; [. . .]’
- d. Dès long-temps les vénitiens ont eu au Kaire des établissemens où ils envoient des selles, des étoffes de soie, des glaces, des *merceries*, etc. (*Frantext*: Comte de Volney, *Voyage en Égypte et en Syrie*, t. 1, 1787, p. 205)
 ‘For a long time, the Venetians had in Cairo establishments where they send saddles, silken cloths, mirrors, *haberdashery*, etc.’

As shown in Figure 8.4 below, over the centuries, there is a tendency for the singular to become predominant, but this mainly holds for locative meanings (LOC), and the frequency of occurrences that have to be interpreted as collection nouns (only SOMNs) fluctuates and becomes negligible from the 20th century on. The marginal occurrences marked as event nouns (EV) in the 19th and 20th centuries

are cases of *mercerie* as a trade or commerce, and the marginal uses as singular object nouns are marked as *SON*.

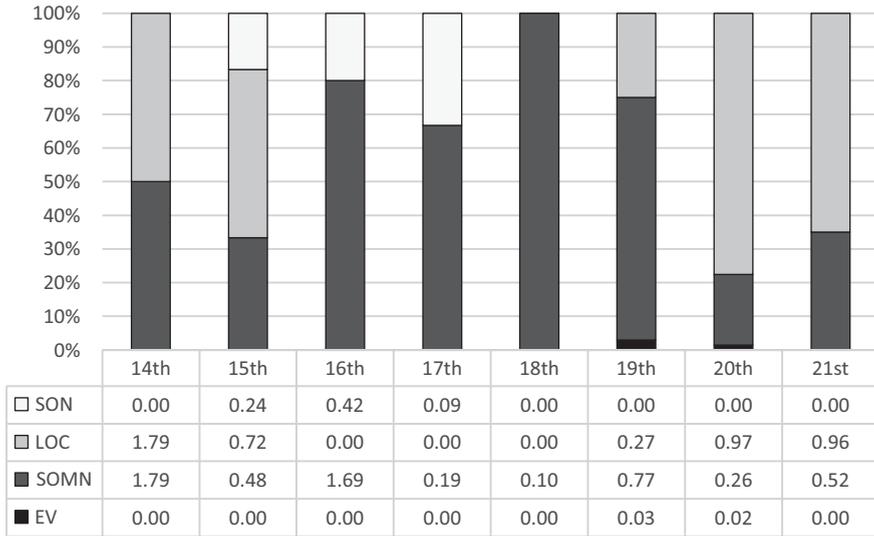


Figure 8.4: Evolution of the nominal types of Fr. *mercerie*_{sG} in *Frantext* (rel. frequencies extrapolated to 1,000,000).

The relative frequencies shown in Figure 8.4 show that the frequency of *mercerie* in the singular is constantly low. The same is true for the inflectional plural: there are e.g. only 3 occurrences of *merceries* in the 21st century (rel. freq. 0.22), all denoting sewing-ware shops. The observation that the proportion of mass vs. ambiguous determiners with *mercerie* in the singular is 14% to 86% in the 21st century would thus be misleading since the overall frequency of *mercerie* as a collection noun is only $7_{\text{abs.}}/0.52_{\text{rel.}}$ in this period. I will consequently leave the analysis of *mercerie* at this rather brief consideration. Summarising, *mercerie* oscillates in the singular between a SOMN and a locative noun from the very beginning. There is a tendency for the singular to become more and more predominant; this, however, also holds for the locative meaning. As in the case of *habillement*, there is a semantic specification from small goods in general to sewing-related products in particular. For *mercerie* there is no evolutionary stage of a CCN documented in the *Frantext* corpus.

Accoutrement: The first attestations of *accoutrement* ‘attire’ in the *Frantext* corpus found at the end of the 15th century. Like *habillement*, it is documented from the beginning in both the singular and the plural, *accoutrement* in the singu-

lar is mostly used as a CCN and the occurrences of *accoutrement* in the plural are predominantly cases of count noun uses referring to single objects. *Accoutrement* hereby confirms what has been assumed for the evolution of collection nouns: instances of plural object nouns reflect a pre-stage of the collection noun. The non-coherent plurality of reference is then bounded to become a CCN. In accordance with the meaning of the derivational base verb MFr. *accoustrer* ‘prepare/adorn’, the meaning of *accoutrement* (both in the singular and the plural) ranges over (items of) war equipment (cf. (82a)), but also adornments (cf. (82b)) and the outer appearance or attire of a person (cf. (82c)). Many plural cases are to a certain extent ambiguous in whether they are referring to individuals or to various kinds of individual. Example (82b) e.g. may also represent a sortal reading – different kinds of adornment. My impression, however, is that most cases primarily refer to a plurality of individuals and not of kinds.

- (82) a. Or, pour parler de *l'accoutrement* du roy, il est assavoir qu'il estoit aussi bien armé en prince de grant renom que jamais homme fut; car il avoit sur luy tout son harnoys complet, beau et riche a merveilles. (*Frantext*: A. de la Vigne, *Le voyage de Naples*, 1495, p. 286)
 ‘Now, to speak of the *equipment/armour* of the king, it should be known that he was as well equipped as a prince with good reputation like never any man was, because he had with him his complete armour, wonderfully beautiful and rich.’
- b. [. . .], sinon que les évesques avoyent quelques *accoustemens* pour estre discernez d'entre les autres Prestres. (*Frantext*: J. Calvin, *Institution de la religion chrestienne, livre quatrième*, 1560, p. 103)
 ‘[. . .] but that the bishops had some *adornments* so that they could be differentiated from the other priests.’
- c. Mais quelle excuse prent-elle d'estre venue voir ma niece en *accoustrement* d'homme? (*Frantext*: O. de Turnèbe, *Les contens*, 1584, p. 115)
 ‘But what excuse had she to come to my niece in the *costume* of a man?’

Syntactically, *accoutrement* shows an evolution similar to *habillement*. With respect to the inflectional number of the occurrences, there is a constant increase in the inflectional singular which, in contrast to *habillement*, does not, however, culminate in an exclusively singular use in Modern French (cf. Figure 8.5). With respect to the use of count and mass determiners, as with *habillement* there is a decrease in count determiners from the 17th century onwards and equally no occurrences of a mass determiner (cf. Figure 8.6).

At first sight, it seems that *accoutrement* mirrors the evolution of *habillement* – *accoutrement*, however, does not become a collection noun which refers

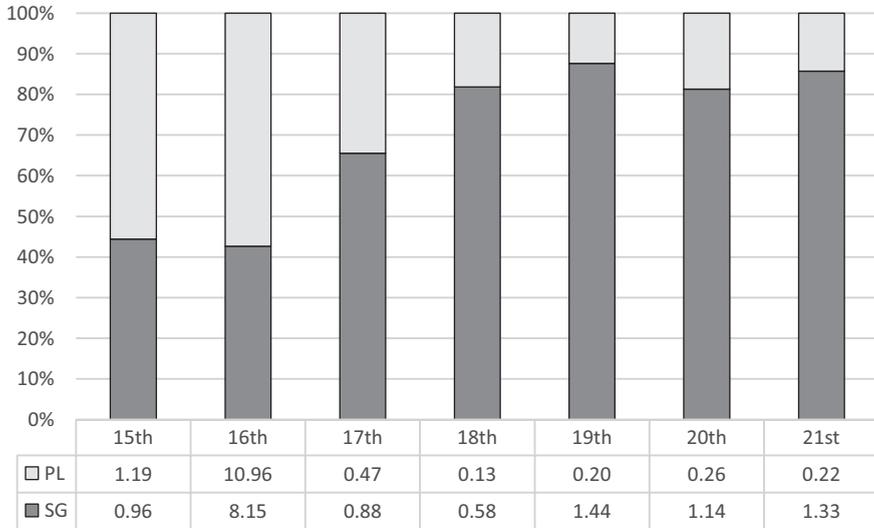


Figure 8.5: Evolution of number inflection of Fr. *accoutrement* in *Frantext* (rel. frequencies extrapolated to 1,000,000).

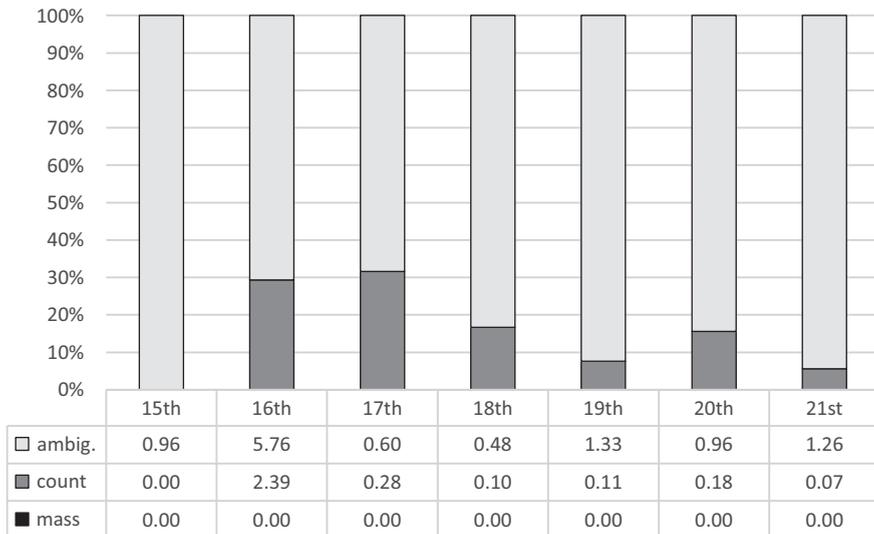


Figure 8.6: Evolution of types of determiner in combination with Fr. *accoutrement*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000).

to a mass superordinate category, but stops at the stage of a CCN whose referents are still to be defined by the context. For this reason, there are many occurrences of *accoutrement* where the constituting elements of the collection in question are listed or described in detail (cf. (83a)); verbs like *compléter* ‘complete’ furthermore support the boundedness of the collection verbalised by *accoutrement* (cf. (83b)).

- (83) a. Adrien n’avait rien cédé quant à son look et se présenta devant le psychiatre avec son *accoutrement* habituel Gothique à souhait. Vêtu de cuir, engoncé dans un lourd manteau en dépit de la température plus que clémente. Le visage maquillé de blanc et les ongles soigneusement vernis de noir, mais tout sourire. (*Frantext*: T. Jonquet, *Ils sont votre épouvante et vous êtes leur crainte*, 2006, p. 104)
 ‘Adrien hadn’t given up on his looks and presented himself to the psychiatrist in his usual Gothic *attire*. He was dressed in leather, crammed into a heavy coat despite the very mild weather. His face was made up in white and his nails carefully polished in black, but he smiled all over his face.’
- b. Un manteau de couleur sombre qu’il avait jeté sur un banc avec son chapeau complétait *l’accoutrement*. (*Frantext*: T. Gautier, *Le capitaine fracasse*, 1863, p. 281)
 ‘A dark coat that he had thrown on a bench with his hat completed the *attire*.’

In Modern French, *accoutrement* is now limited to the pejorative denotation of a slightly eccentric costume or outfit. The evolution of the nominal types of *accoutrement* in the singular are summarised in Figure 8.7 below. As shown, the proportion of CCN interpretations increases over time while uses of singular object nouns (SONs) as well as ambiguous cases decrease. Occurrences in the plural vary over the centuries between interpretations as plural object nouns and CCNs by not predominantly and uniformly favouring one or the other option.

Summarising the evolution of *accoutrement*, it shows stages of development very similar to those of *habillement*. It originates as a noun denoting either single items of equipment or adornment or the bounded equipment or attire of a person. Syntactically, there is in both cases a clear increase in the singular inflection and a decrease in count determiners like the indefinite article. In contrast to *habillement*, however, *accoutrement* does not become a SOMN, but stops at this earlier stage.

Équipement: In the synchrony of present-day French, *accoutrement* and *équipement* share the property of being deverbal CCNs. Apart from this though, they do not have much in common. The CCN *équipement* ‘equipment’ is attested in the *Frantext* corpus from the beginning of the 18th century. Similar to the der-

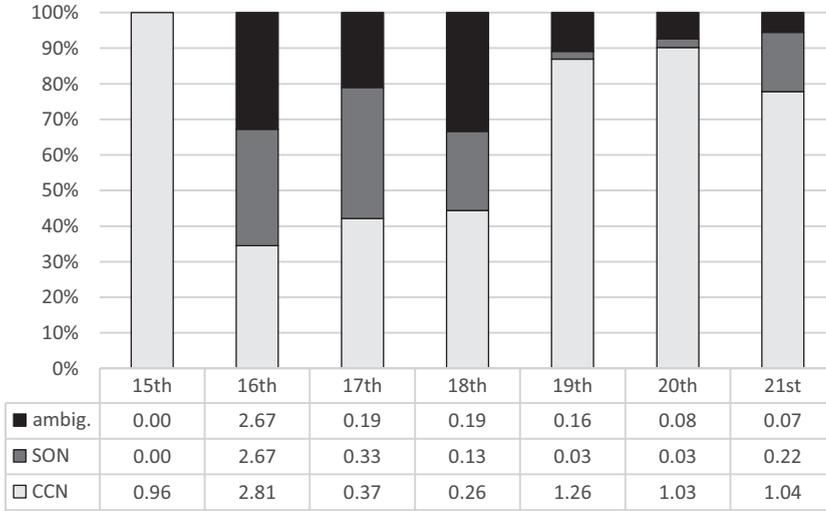


Figure 8.7: Evolution of the nominal types of Fr. *accoutrement*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000).

ivationally equal *habillement* and *accoutrement*, there are in the beginnings of the evolution of *équipement* contexts where the noun has to be interpreted as a CCN (cf. (84a)), but in contrast to the other two derivates on *-ment*, there are also eventive meanings attested for *équipement* (cf. (84b)).

- (84) a. [. . .] et Ali Baba, qui en compta quarante, à leur mine et à leur *équipement*, ne douta pas qu'ils ne fussent des voleurs. (*Frantext*: Anonymous, *Mille et une nuits*, t. 2, 1715, p. 414)
 '[. . .] and Ali Baba, who counted forty of them, at their mine and their *equipment*, did not doubt that they were thieves.'
- b. Les deux nations française et batave se serviront également du port et du bassin de Flessingue pour la construction, la réparation et *l'équipement* de leurs vaisseaux. (*Frantext*: E. Sieyès, *Œuvres*, t. 3, 1799, p. 5)
 'The two nations, France and the Netherlands, will use both the port and the basin of Vlissingen for the construction, repair and *equipping* of their ships.'

The two derivationally equivalent collection nouns *habillement* and *accoutrement* originated as a variety of different aspectual types and in both cases, there was a clear tendency of limitation to one single type – a SOMN in the former and a CCN in the latter case. There is no such evolution in the case of *équipement*. Until the present, *équipement* has denoted a spatio-temporally bounded collection of

things that serve to equip somebody or something, items of equipment, the event of equipping and also equipment in general, conceived of as an indefinite mass. The CCN interpretation is illustrated in (85a) below, where the adjective *petit* ‘small’ cannot refer to the plurality of items making up the equipment, but only to the quantity of them. It is thus an example similar to Sp. *gentío grande* ‘big crowd’ mentioned in chap. 3.1, where the adjective *grande* ‘big/tall’ cannot access the individual people, but only the collection itself. The set profiling of the collection is supported by the non-distributive quantifier *tout son* ‘all her’. It is interesting to note that although *équipement* clearly denotes a bounded collection in this context, it can be combined with a verb like *rassembler*, which accesses internal plurality. This is therefore an illustrative example of the fewer restrictions these kinds of verbs put upon the collections nouns they are combined with, just like *la familia se reunió* ‘the family assembled’ mentioned in chap. 3.1.2. In addition to these clear CCN interpretations, *équipement* may also denote single items of equipment (cf. (85b)) and to some extent also the action of equipping somebody or something (cf. (85c)). *Armement* ‘arming/arms’ in this latter example may also refer to the amour as a collection noun, but the binominal construction on the base of *efforts* ‘efforts’ rather points to an eventive interpretation here. Especially in the case of administrative tracts found in the corpus, there also is the possibility of *équipement* to be used as a SOMN. Example (85d) illustrates this use by means of the now lexicalised expression of *équipement collectif* ‘public facilities’. These kinds of occurrence were categorised as OMNs since the constituting elements of the collection are linked not by some kind of contiguity, but by a similar function. With this interpretation, they are kinds of equipment and not part of the equipment. These cases are, however, not prototypical for this nominal type since the exact extension may still be left unspecified. That is to say that from a referential point of view, the items of equipment are only determined by para-relations, i.e. expectation relations. With this the occurrences of *équipement* as a SOMN are referentially still very close to the status of a CCN (cf. chap. 7.2.2).

- (85) a. Sylvie, désespérée, n’avait pas eu le temps de rassembler tout son *petit équipement féminin*, épars sur la table et sur la banquette, que les deux hommes avaient disparu. (*Frantext*: H. Vincenot, *Le Pape des escargots*, 1972, p. 177)
 ‘Sylvie, distraught, did not have the time to gather all of her *few female belongings*, scattered over the table and the bank, that the two men had vanished.’
- b. Ils peuvent aller un peu plus au cinéma, ils peuvent avoir quelques *équipements ménagers supplémentaires*, [. . .] (*Frantext*: P. Mendès-France, *Œuvres complètes*, t. 6, *Une vision du monde*. 1974–1982, 1990, p. 514)

‘They can go to the cinema a little more often, they can have some additional domestic *appliances*, [. . .]’

- c. C’est de savoir si les efforts d’armement et d’équipement sont présentés comme favorables ou discutables. (*Frantext*: A.-M. Garat, *Dans la main du diable*, 2006, p. 755)

‘We have to know whether the arming and *equipping* efforts are presented as favourable or as debatable.’

- d. En matière d’équipement collectif, on pourrait difficilement admettre que l’action des pouvoirs publics a trouvé sa forme définitive. (*Frantext*: L.-M. Jocard, *Le tourisme et l’action de l’état*, 1966, p. 194)

‘Regarding the *public facilities*, we can hardly admit that the actions of the public authorities have found their final form.’

In the first two centuries of evolution, the variation between different nominal types is found in *équipement* for both the inflectional singular and plural, while the plural *équipements* is mostly found denoting individual items of equipment, but also industrial installations in the 20th and 21st centuries. In the case of the singular, there is no perceptible evolution towards the predominance of one or the other type. Figure 8.8 summarises the analysis regarding the nominal types of *équipement* for the inflectional singular (EV = eventive, SON = singular object noun).

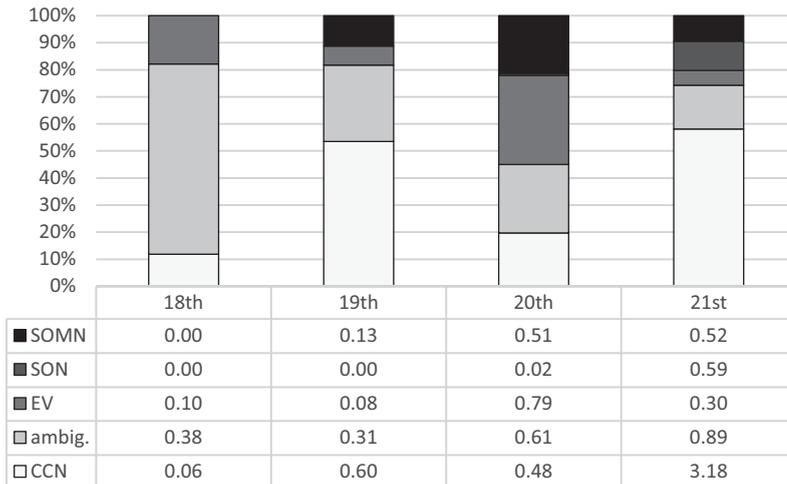


Figure 8.8: Evolution of the nominal types of Fr. *équipement*_{sg} in *Frantext* (rel. frequencies extrapolated to 1,000,000).

Summarising the analysis of *équipement*, it again evidences another kind of development. While an unambiguous evolution towards a SOMN was shown for *habillement* and a clear pathway towards a CCN in the case of *accoutrement*, *équipement* has been a flexibly used noun from its beginnings until very recently.

Fringues: Fr. *fringues* ‘clothes’ is only attested from the 20th century; it then exclusively occurs with plural inflection and has to be interpreted mostly as a POMN. There are some very marginal cases of *fringue* in the singular in the 21st century (abs. freq.: 3) which either refer to single pieces of clothing (cf. (86a)) or clothing as an indefinite mass (cf. (86b)). There are equally some cases where *fringues* in the plural is combined with a distributive determiner, but these are mostly identifying adjectives like *les mêmes* ‘the same’, *les autres* ‘the other’ or *certain* ‘certain’ – the case in (86c), where *fringues* is combined with a numeral, is unique in the corpus. The POMN is combined with ambiguous determiners in over 97% of occurrences in this century and the last.

- (86) a. Ange m’a offert le blouson de cuir qu’il avait dépouillé à l’un d’entre eux. “Je mets plus ce genre de *fringue*”, m’a dit-il. (*Frantext*: S. Osmont, *Éléments incontrôlés*, 2012, p. 96)
 ‘Ange gave me the leather jacket that he had taken from one of them. “I don’t wear this kind of *piece of clothing* anymore”, he said to me.’
- b. mais comment le retrouver, lui, dans le royaume du fast-food et de la *fringue* que ce quartier est devenu? (*Frantext*: M. Audin, *Une vie brève*, 2012, p. 137)
 ‘but how can he be found, him, in the kingdom of fast-food and of *cheap clothing* that this district has become?’
- c. [. . .] je me suis fait arnaquer moi aussi, bien comme il faut. Coup sur coup, mes deux *fringues* préférées, comme un cave. (*Frantext*: Bayon, *Le Lycéen*, 1987, p. 373)
 ‘[. . .] I have been ripped off as well, just as it should be. One after the other, my two favourite *clothes*, like a cellar.’

Figure 8.9 below summarises the scant data available for *fringues*, where the very marginal object noun uses are marked with *ON* in the graph.

Summarising this analysis of *fringues*, there is not much data available in the *Frantext* corpus from which to draw conclusions. One may, however, conclude that *fringues* probably does not match the pattern of the other deverbal collection nouns discussed in this chapter. It is directly first attested in the plural as a POMN and, at least in the two most recent centuries, there has been no apparent evolution of *fringues*, neither syntactically, nor semantically. This is in line with

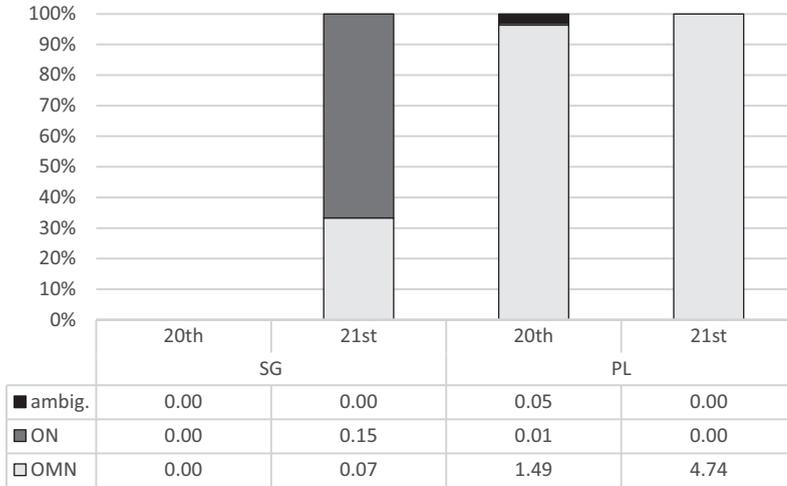


Figure 8.9: Evolution of number inflection and of the nominal types of Fr. *fringues* in *Frantext* (rel. frequencies extrapolated to 1,000,000).

the observation of Mihatsch (2006, 131) that pluralisation in the same semantic domain may also be influenced by analogy. Many colloquial expressions for clothing in French represent pluralia tantum or are at least mostly used in the plural. This holds for e.g. *fringues*, *frusques*, *sapes*, *nippes* and *hardes*, all referring derogatorily in colloquial speech to clothing. German also has a number of derogatory terms for clothes or small things like *Klamotten* ‘clothes’. These are often also regionally marked like *Plörren* (Ruhr-area), *Plünnen* (Northern Germany), *Plünsen*, *Prütten*, *Prötten* or *Piselotten* (all four Sauerland-region) which refer to small, negligible stuff.⁷¹ It may be that two effects have reinforced each other here: on the one hand, there are possible semantic effects. The absence of a CCN and a SOMN stage in these cases may be a manifestation of the expression of a non-coherent collection of unimportant things. If this is true, the plural is not a symptom of the growing similarity of referents (cf. Mihatsch 2006, 128), but rather of a neglecting of the functional contiguity of the referents. On the other hand, since there is a remarkable number of quasi-synonyms in the French and German examples, it may be that formal analogy may have supported the tendency towards pluralisation. Further research is necessary to examine this hypothesis further.

⁷¹ I am indebted to my dear friends and colleagues who shared with me this collection of regional terms.

Vêtement: The last collection noun in this section on functional artefact nouns is *vêtement* ‘clothing/piece of clothing’. It is treated along with *habillement*, *accoutrement* and *équipement* as a derived artefact noun, but strictly speaking it should be discussed separately. Fr. *vêtement* is a direct continuation of Lat. VESTĪMENTUM, and the deverbal derivation thus already happened before the coming into being of the Romance languages. In Latin, *vestimentum* in the singular mostly refers to single pieces of clothing, the toga in particular (cf. (87a)); the inflectional plural *vestimenta* then denotes a plurality of clothes (cf. (87b)). Note that the Latin neuter plural is often not to be interpreted as a distributive plural, but is more a collective plural (cf. Schön 1971, 56, cf. chap. 3.2).

- (87) a. Nam uti quisque domum aut villam, postremo vas aut *vestimentum* aliquoius concupiverat, dabat operam, uti is in proscriptorum numero esset. (Sallust: *De coniuratione Catilinae*, 61 b.c)
 ‘For if anyone desired someone’s house or estate, even vessel or garment, he endeavoured to have that person put on the list of outlaws.’
- b. ego vos in duas iam pelles coniciam vinctos que loris inter *vestimenta* pro sarcinis habebō, [. . .]. (Petronius: *Satyrica*, 1st cent. a.c.)
 ‘I will put you in two leather bags, tie straps around them and put them between my *clothes* as luggage, [. . .]’

In the early attestations in *Frantext*, *vêtement* patterns mainly with its Latin etymon: the inflectional singular refers to one single garment and the plural to more than one (cf. (88a)). There are, however, also some marginal first attestations of a mass superordinate use. In (88b), the verb *atourner* ‘equip/provide’ clearly implies more than one piece of clothing.

- (88) a. Si enveiad á Técúe é de la fist venir une sage femme, si li dist: “Fâi semblant de duléir é de plur, si t’afúble de *vestment de plur* si cume femme ki lunges ait pluréé pur mort. (*Frantext*: Anonyme, *Li quatre livre des reis*, 1175, p. 83)
 ‘So he sent men to Tecue and from there, he had a wise women come, and so he said: “I make you look like a mourning woman, so I dress you in a *mourning garment/robe* as if you were a woman who has been crying for a dead man for a long time.”’
- b. “Sire, fet ele, mout grant honte sera a vos, plus qu’a autrui, se cist sires an mainne o lui vostre niece si povremant atornee de *vestment*.” (*Frantext*: Chrétien de Troyes, *Erec*, 1170, p. 2a)
 ‘Sir, she said, shame should be on you, more than on anyone else, if this sir brings with him your niece so poorly provided with *clothing*.’

These two aspectual interpretations of *vêtement* remain in use over the whole period of time under investigation, and there is no considerable semantic change as was the case with *habillement* e.g. – *vêtement* always refers to the (outer) pieces of clothing without shoes, accessories and underwear. In the corpus this is attested by numerous occurrences of *vêtement* in coordination with – and thus in opposition to – *atournement* ‘adornment’, *chaussure* ‘shoes’, *linge* ‘underwear’ or *chemise* ‘vest’, and *vêtement* in anaphoric reference to *manteau* ‘coat’, *robe* ‘garment’ or *redingote* ‘coat’. Occurrences of CCNs are only very marginally attested for *vêtement* and these are mainly ambiguous between the spatio-temporally contiguous collection of pieces of clothing and clothing conceived of more generally or as a single piece of clothing. Example (89) illustrates the former possibility, whereby *vêtement* may refer either to the costume of an actor or to clothing in the context of theatre in general:

- (89) On doit à Talma la perfection de la tenue de l'acteur. Mais la vérité du théâtre et le rigorisme du *vêtement* sont-ils aussi nécessaires à l'art qu'on le suppose? (*Frantext*: F.-R. de Chateaubriand, *Mémoires d'outre-tombe*, t. 2, 1848, p. 38)
- ‘We owe to Talma the perfection of the costume of the actor. But are the truth of the theater and the rigour of *the costume/the clothing* as necessary as we assume them to be?’

The steady fluctuation of aspectual types of *vêtement* is summarised in Figure 8.10 below for the inflectional singular (singular object nouns are again indicated as SON), the plural occurrences uniformly represent cases of plural object nouns. The absence of any tendency toward a certain aspectual type in the case of *vêtement* is reflected by the types of determiner it is combined with. As depicted in Figure 8.11 below, there is a steady proportion of about one-third to two-thirds of count vs. ambiguous determiners of which the former mainly represent the clear cases of singular object nouns shown earlier in Figure 8.10. Most of these count determiners are instances of the indefinite article, attested from the very first occurrences in the corpus, but other distributive quantifiers like *aucun vêtement* ‘no (single) piece of clothing’ and identifying adjectives like *son dernier vêtement* ‘his last piece of clothing’ are also well documented. As in the case of the other nouns examined in this section, there are no attestations of the partitive article clearly marking mass syntax, and neither are there occurrences of a classifier construction – which is not necessary in the case of *vêtement* given its polysemy.

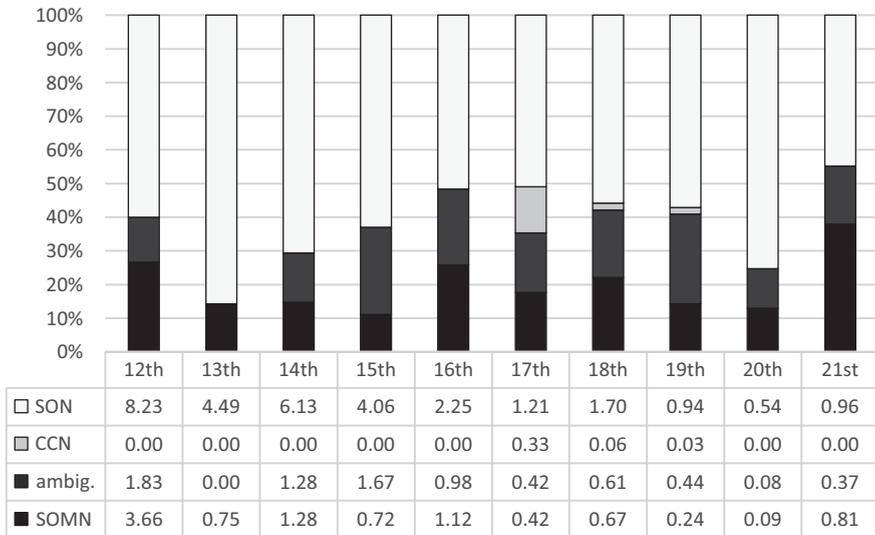


Figure 8.10: Evolution of the nominal types of Fr. *vêtements*₅₆ in *Frantext* (rel. frequencies extrapolated to 1,000,000).

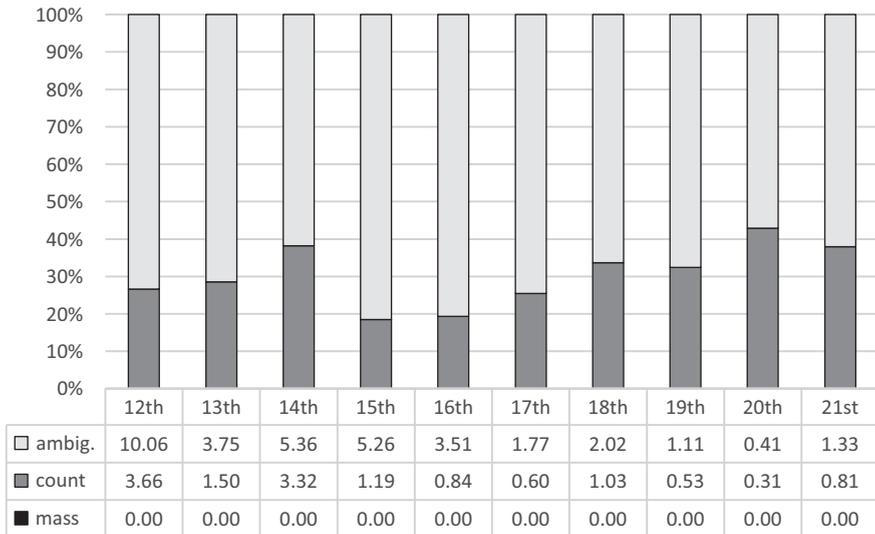


Figure 8.11: Evolution of types of determiner in combination with Fr. *vêtements*₅₆ in *Frantext* (rel. frequencies extrapolated to 1,000,000).

Summarising the analysis on *vêtement* so far, it represents a direct continuation of its Latin etymon *VESTĪMENTUM* meaning first of all ‘piece of clothing’ and to a lesser extent also ‘clothing’. The corpus analysis did not show any considerable changes in the semantics or in the syntactic restrictions with regard to its countability. *Vêtement* as an indefinite mass of clothing is already present in Old French with no intermediate CCN stage attested. The only apparent change in the linguistic characteristics of *vêtement* is the proportion between the inflectional singular and plural. As mentioned above, both forms are present from the 12th century onwards and remain present over the centuries analysed. The singular inflection, however, shows a slight tendency to decrease over time. While the proportion between the singular and plural was more or less equal in the 12th century, there is a clear predominance of the plural inflection in recent texts from the 21st century:

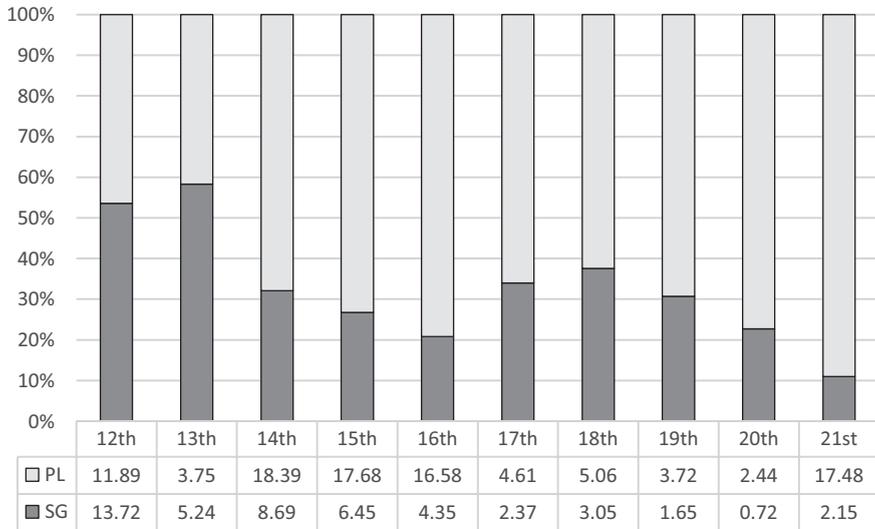


Figure 8.12: Evolution of number inflection of Fr. *vêtement* in *Frantext* (rel. frequencies extrapolated to 1,000,000).

In light of these figures, *vêtement* can be assumed to be on its way to becoming a POMN. This assumption can be supported by a constantly low proportion of distributive quantifiers in combination with the inflectional plural, ranging from zero to 5% over the centuries. However, there is also the tendency of superordinates in the non-biological domain to occur more often in the plural, in contrast

to basic level nouns which mostly occur in the singular (cf. Wisniewski/Murphy 1989; Mihatsch 2006, 144). Supposing that *vêtement* is nearly a POMN lacking an equivalent inflectional singular form, this can be taken as another illustrative example of the unidirectional lexicalisation pathway of collection nouns as assumed by Mihatsch (2016). There is, however, an inconsistency in the chronology: the plural inflection did not originate in the SOMN, but was present from the very beginning. Fr. *vêtement* is thus not equivalent to *habillement*, *accoutrement* or *équipement* and therefore does not represent an illustrative case of a possible lexicalisation path of collection nouns, but should be addressed on its own terms. It may be that its special status is due to the lack of an overt mass-count distinction in Latin and also in Old French (cf. chap. 1.2.2). Similar to Brazilian Portuguese (cf. chap. 4.2), Latin did not provide the necessary syntactic structures for an overt distinction between mass and count superordinates. As such, *vestimentum* is continued into French mainly as a count noun referring to individual pieces of clothing and only marginally as a SOMN. The strong degree of similarity of the referents of OFr. *vestment* probably hindered an increase in occurrences where it could be conceived of as an indefinite mass: as outlined in chap. 3.1, mass superordinates are a linguistic compromise between a plurality of heterogeneous referents and their conceptualisation as one collection, between class inclusion and group membership. It would not have been linguistically economic to (re)introduce the heterogeneity necessary for a SOMN. There is still the possibility that the assumed stages of development of a collection noun had already happened in early Latin; these paths are, however, not traceable in the context of this present work. At least in the evolution of French, *vêtement* is consequently next to or ahead of the assumed lexicalisation path of collection nouns.

Summarising these remarks about the derived artefact collection nouns examined here, the following may be stated: although the nouns analysed share a number of semantic and morphological features, they have all evolved along different pathways. For *habillement* and *accoutrement*, it they could be said to represent illustrative examples of the assumed evolution of collection nouns, with *habillement* now at the stage of a SOMN and *accoutrement* of a CCN. However, *équipement* and *vêtement* did not follow the expected path, either because their history is still too young – the future development of *équipement* can only be guessed – or they are already evolving ahead of the assumed stages, as has been supposed for *vêtement*. Finally, *mercerie* and *fringues* had to be neglected for this present analysis because of their low frequencies of occurrence as collection nouns.

8.3.2 Collection nouns based on one shared property: *mobilier*, *gens*

The two collection nouns to be discussed in the following are classified as being based on one shared property. This classification should not, however, be taken too strictly. As a typical example, Germ. *Mannschaft* ‘team’ was described earlier as denoting first the state of being in relation to an overlord and then, via a metonymic shift, the collection of men (and later also women) which share this property. The two French collection nouns *mobilier* and *gens* show some differences to this prototypical example, from the point of view of derivation and also of semantic change. As has been elaborated above, *mobilier* represents a case of lexical absorption or ellipsis where the meaning of the whole complex word or construction in examples like *effets mobiliers* or *biens mobiliers* was transferred to the adjective and the noun was eliminated. Fr. *mobilier* thus designates, in contrast to Engl. *furniture* but similar to other Romance equivalents like Pt. *mobília*, not a functional artefact noun but a collection of artefacts bound by a particular property. One may say that, in its origins, Fr. *mobilier* was a stage-level artefact noun since the items belonging to this collection only became *mobilier* when they became part of the movable property of a person. Just as Engl. *mail*, a letter and a chair are by default not *mail* and *mobilier*, but only become such when taking part in a specific event associated with the collection noun. In comparison to Engl. *mail*, my impression regarding *mobilier* is that the items of the collection are linked more by the unifying feature of being movable property than by the associated event of being bought. In the case of Fr. *gens*, the shared property of the items of the collection lies in their being of the same origin. This meaning is present to a lesser extent in French, but in Latin, *gēns* designated the family, a tribe or a nation. In its further development, Fr. *gens* first denoted the retinue of a monarch, i.e. people in relation to a central attraction point. This relation is loosened more and more and in Modern French *gens* denotes people in general. From a derivational point of view, *gens* is not a complex, but a simplex word, thus different from Germ. *Mannschaft* and Fr. *mobilier*. These elaborations show that *mobilier* and *gens* may well be classified as collections originating on the basis of one shared property, but they are not prototypical cases of this category.

Given the lack of an associated event – the starting point of the assumed lexicalisation path of collection nouns – in both cases, it is difficult to formulate a consistent hypothesis with regard to the semantic and morpho-syntactic evolution of these two examples. One may nevertheless suppose that, even if their origins cannot be classified under the pattern of the functional artefact collection nouns, they may follow the path from the point of the CCN where the associated event is no longer in focus. The synchronic analyses presented in chap. 4 showed that *gens* is a POMN in Modern French and *mobilier* may be somewhere

in between a CCN and a SOMN. I consequently assume the change from a CCN to a SOMN and then to a POMN for *gens*; for *mobilier*, I assume an attested status as an CCN and some slight tendencies towards a SOMN in recent texts.

Mobilier: The first occurrences of *mobilier* as a noun appear in the *Frantext* corpus in the middle of the 18th century. These first attestations are all in the singular and semantically still very close to the etymologically original meaning of movable property. Interestingly, although *mobilier* stems from *biens mobiliers* ‘movable property’, these first attestations in the corpus do not occur in administrative texts, but mostly in personal correspondence or travel diaries but, as will be shown later, the link to jurisdiction has not, however, been completely lost. In the 18th century, the collection noun is semantically and syntactically very clearly to be categorised as a CCN: it is often combined with count determiners like the indefinite article; the referents mentioned in the context are heterogeneous and not limited to wooden furniture; adjectives like *immense* ‘large’ modify the size of the collection and not the constituting entities; and *mobilier* is often combined with verbs like *consister de* ‘consist of’, pointing to a meronymic relation between the collection and its entities. All these linguistic characteristics point to a heterogeneous collection with firm outer boundaries, the collection noun is countable, and the constituting entities are not accessible. Examples (90a) and (90b) illustrate this, where *mobilier* may be best translated with ‘(movable) property’ here. There are, however, some very marginal cases like (90c) where such a clear categorisation as a CCN is not possible. In this context, *mobilier* is not a spatio-temporally bounded collection, but an indefinite mass. The context does not give any information on the exact kind of referents *mobilier* refers to, but one may imagine perhaps furnishings or pieces of furniture. This example was thus categorised as an SOMN given the lack of boundedness and a supposed kind-of relation between the constituting entities and the collection itself. This is not, however, a categorisation based on firm criteria like the absence or presence of a certain kind of determiner, but rather on an impression of the contextual information given.

- (90) a. Une fille d’opéra qui vient de décéder, laisse *un mobilier immense*, une somme d’argent considérable. (*Frantext*: L.-S. Mercier: *Tableau de Paris*, t. 7, 1783, p. 10)
 ‘An opera singer, who just passed away, leaves *a large property* (i.e. *possession*), a large sum of money.’
- b. [. . .] tout mon *mobilier*, consistant en linge, meubles et argent comptant, qui se trouverait en ma possession au jour de mon décès en pays étrangers, et en outre la somme de douze mille livres. (*Frantext*: G. Sénac de Meilhan: *L’émigré*, 1797, p. 1899)

‘[. . .], all my *possessions*, consisting of linen, pieces of furniture and cash, which would be in my possession the day of my death in foreign countries, and in addition, the sum of twelve thousand pounds.’

- c. On pourroit néanmoins en excepter les sommes immenses employées en *mobilier* de pure fantaisie, qui n’a de prix réel en quelque sorte que par la mode; [. . .] (*Frantext*: V. marquis de Mirabeau: *L’Ami des hommes ou Traité de la population*, t. 1, 1755, p. 33)

‘One might nevertheless exclude the large sums of money invested in *furniture/furnishings* of pure fantasy, which has/have no real price but in fashion; [. . .]’

Over the period of the next three centuries until present-day French, the predominance of the singular inflection remains present, as can be seen in the following Figure 8.13:

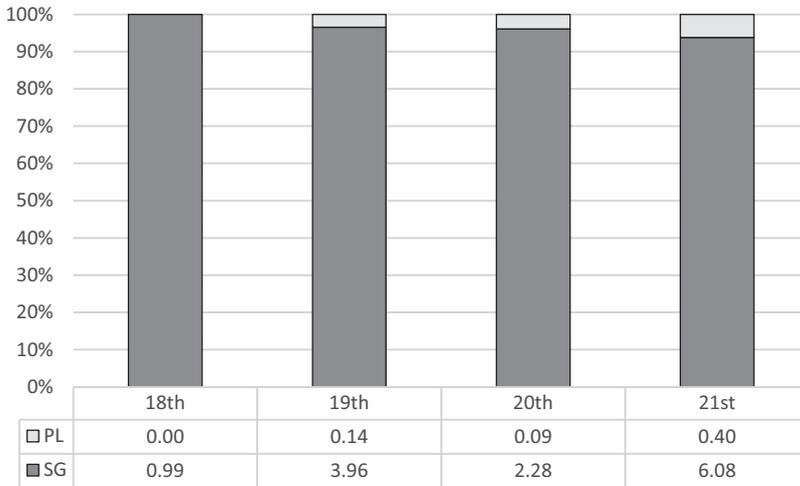


Figure 8.13: Evolution of number inflection of Fr. *mobilier* in *Frantext* (rel. frequencies extrapolated to 1,000,000).

The very few cases of *mobilier* in the plural oscillate between either an interpretation as single pieces of furniture or as sets of furniture/furnishings. The context often does not provide enough information to unambiguously classify the aspectual type. In example (91a), it is e.g. not clear whether this is referring to single pieces of furniture that are soaked from the rain, or rather complete sets in the single rooms of the building. In example (91b), however, the adjective *complet*

‘complete’ and in example (91c) the contrast with *meubles* ‘pieces of furniture’ tend to point to a collective interpretation.

- (91) a. La plupart des fenêtres, dépourvues de leurs vitres, laissaient entrer la pluie qui lézardait les plafonds, tachait les murailles, *les mobiliers*. (*Frantext*: M. Aymé, *Nouvelles complètes*, 2002, p. 110)
 ‘Most of the windows, lacking their windowpanes, let the rain in which cracked the ceilings, stained the walls and *the pieces of furniture/the furnishings*.’
- b. Caledonian Market vend des jacinthes d’occasion et des fragments d’ispahans, des clubs de golf rouillés, *des mobiliers complets*, rêves de nomades, ou des machines à écrire à touches jaunes, comme de vieux dentiers. (*Frantext*: P. Morand, *Londres*, 1933, p. 268)
 ‘Caledonian Market sells second-hand hyacinths, fragments of Isfahan carpets, rusty golf clubs, *complete sets of furnishings*, nomads’ dreams, or type writers with yellow keys, like old dentures.’
- c. Là s’entassaient des *mobiliers* de toutes les époques, particulièrement des meubles de salon somptueux. (*Frantext*: A. Dhôtel, *Le Pays où l’on n’arrive jamais*, 1955, p. 220)
 ‘*Sets of furniture* of all epochs piled up there, especially sumptuous pieces of furniture for lounges.’

The singular occurrences of *meublier* show a constant homogenisation of the referents from movable property to furnishings to wooden furniture, excluding decorations etc. While *meublier* is still contrasted with *immeubles* in the *Code Civil Français* (1804) – thus movable vs. immovable property – it increasingly occurs with modifying prepositional phrases like *de ce salon* ‘of this lounge’ or *de cette pièce* ‘of this room’ indicating a semantic change towards the furnishings of a room (cf. (92a)). It often still occurs with verbs indicating a meronymic relation between a collection and its entities until the 20th century, pointing to an interpretation as a CCN. Parallel to this, there is also an increasing number of occurrences where *meublier* is modified with adjectives and prepositional phrases that only allow for an interpretation as wooden furniture (e.g. *meublier en acajou* ‘mahogany furniture’) and combined with verbs or adjectives which can only refer to the constituting entities of a collection. An example of the latter case is illustrated in (92b), where *meublier* is modified by the stubbornly distributive predicates *rondissant* ‘rounded’ and *anguleux* ‘angular’. These occurrences point on the one hand to a homogenisation of the referents and on the other to a weakening of their outer boundedness. Most importantly, there are first attestations of *meublier* in combination with a mass-marking partitive article in the first half of

the 20th century.⁷² This is illustrated in (92c), where *du mobilier* refers to an indefinite mass of wooden furniture.

- (92) a. Un tapis de feutre aux couleurs fanées, six chaises et un canapé revêtus de housse, achevaient le *mobilier* de cette chambre, [. . .] (*Frantext*: P. Bourget, *Mensonges*, 1887, p. 104)
 ‘A felt carpet in faded colours, six chairs and a sofa draped with a cover, completed the *furnishings* of this room, [. . .]’
- b. Ce pays, qui a eu le coquet et *rondissant mobilier* de paresse du XVIIIe siècle, il est sous la menace de ce dur et *anguleux mobilier*, qui semble fait pour les membres frustes d’une humanité des cavernes et des lacustres? (*Frantext*: E. de Goncourt, *Journal: mémoires de la vie littéraire*, t. 4, 1896, p. 893)
 ‘This country, which had the pretty and *round lazy furniture* of the 18th century, is under threat of this hard and *angular furniture*, which seems to be made for the frustrated members of a cave- and lake-dwelling mankind?’
- c. Dominique l’attendait, elle alla vers lui et ils s’assirent face à face, dans du *mobilier* confortable. (*Frantext*: R. Queneau, *Loin de Rueil*, 1944, p. 145)
 ‘Dominique was waiting for him, she went towards him and they sat face to face on comfortable *furniture*.’

Summarising this mostly qualitative analysis, there is a slight, but steady increase in occurrences, where *mobilier* has to be interpreted as a SOMN. More importantly, there is, however, also a marked increase in ambiguous cases, where a clear classification as either a CCN or a SOMN is not possible. In addition, there is a very marginal but steady proportion of cases where *mobilier* has to be interpreted as referring to a single piece of furniture (SON). These impressions are summarised quantitatively in Figure 8.14 below.

72 There is also another, earlier case in the *Code Civil Français* (1804), which should however be interpreted as a case of a contextually driven massification of a CCN and not as a *real* mass noun:

- (i.) En cas d’insuffisance, le donataire peut se dispenser de rapporter du numéraire, en abandonnant, jusqu’à due concurrence, du *mobilier*, et à défaut de *mobilier*, des immeubles de la succession. (*Frantext*: Collective authorship, *Code Civil Français*, 1804, p. 157)
 ‘In the event of insufficiency, the donee may dispense with bringing in cash, abandoning, up to the appropriate amount, *movable property* and, in the absence of *movable property*, the immovable property of the estate.’

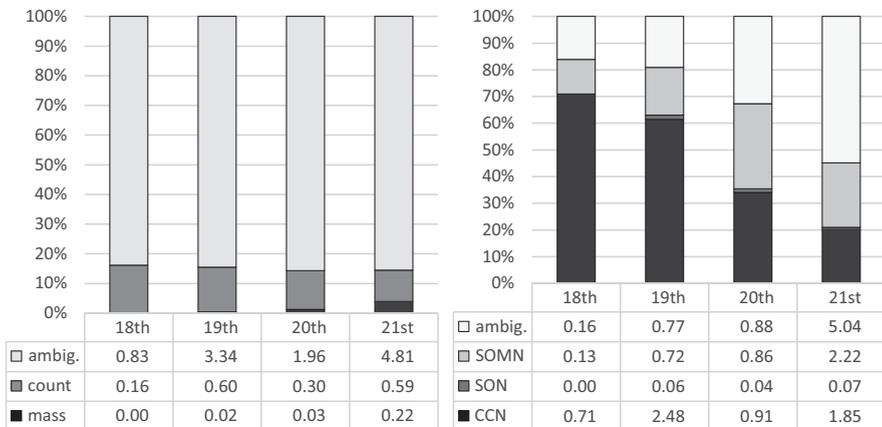


Figure 8.14: Evolution of types of determiner in combination with Fr. *mobilier*_{SG} and evolution of the nominal types of Fr. *mobilier*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000).

The diachronic evolution of *mobilier* with regard to the hypothesis of a unidirectional lexicalisation path of collection nouns may be summarised as follows: Fr. *mobilier* originates via lexical absorption as a CCN meaning ‘movable property’. Its referents are thus heterogeneous, it is countable and the referents are not linguistically accessible. Over time, the referents become more and more homogeneous, the outer boundaries of the collection weaken and there are first attestations of the partitive article. Given the large number of ambiguous cases, the analysis of the diachronic evolution of *mobilier* consequently confirms the hypothesis formulated above: it follows the pathway of collection nouns from the CCN stage and further, and lies somewhere between a CCN and a SOMN in present-day French.

Gens: This collection noun is attested in the whole period of time covered by the *Frantext* corpus, that is to say from the 12th to the 21st centuries. In Old French, the semantic-syntactic characteristics of *gens* are still very close to its Latin etymon: there is a predominance of the singular inflection and it mainly refers to specifically delimited collections like tribes, the retinue of a monarch, or armed forces. From the 12th century onwards, there is then a relatively rapid decrease in these singular occurrences until they are only marginally attested in the 16th century and completely absent from the 17th century on.⁷³

⁷³ The singular form *gent* still survives very marginally in its very original meaning of ‘nation’ or ‘people’, thus the etymological meaning. However, these cases have a too low frequency to still appear in my random sample (see Ising 2019, 326–332 for a similar corpus analysis on this noun and which also takes into consideration these marginal cases).

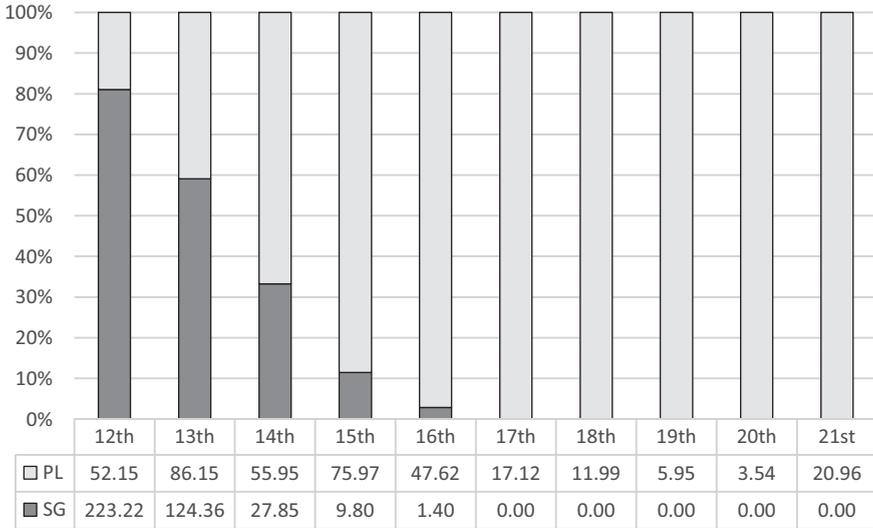


Figure 8.15: Evolution of number inflection of Fr. *gens* in *Frantext* (rel. frequencies extrapolated to 1,000,000).

In considering only the evolution of number inflection, *gens* can be assumed to reflect the evolution of a CCN to a POMN with a possible intermediate stage of a SOMN (cf. Figure 8.15). To examine this assumption, I will first analyse the occurrences of *gens* in the singular and then those in the plural. As has been mentioned above, most of the singular occurrences of *gens* denote bounded collections and should therefore be categorised as CCNs. This is most obvious in the modifiers that *gent* is combined with. On the one hand, there are many contexts in the corpus where two armies or two people are contrasted. Often it is the *gent mescreant*, *sarrasine*, *barbarine*, *estrange*, *païen* ‘unbelieving, Saracen, barbarous, foreign, pagan people’, thus the heathen enemies of the Christian monarchs praised in the medieval heroic lays. As indicated by some of the adjectives, *gent* is often, but not always, marked as having feminine gender, just like the Latin etymon from which it derives. In Modern French, *gens* may well be combined with both feminine and masculine adjectives (cf. Grevisse/Goosse 2016, §490), a systematic study on the evolution of the gender of *gens* cannot, however, easily be done since the gender of modifiers varies depending on their position. On the other hand, there are several occurrences where *gent* is combined with adjectives like *grand* which in particular contexts may only refer to the quantity of people making up a collection and not to the quality of being tall persons. This is illustrated in (93a) and (93b) below, where in the latter example the set-profiling of the CCN is still reinforced by the use of the indefinite article. In contrast to every other

collection noun examined here, there is a considerable number of constructions *ad sensum* in combination with *gent*. In (93a) the plural inflection of the verbs *naiger* ‘to navigate’ and *singler* ‘to sail’ may only be interpreted as anaphoric reference, but in (93c) *gent* is combined directly with a verb in the plural (*apelers* ‘to call’) with no space between subject and predicate. This is remarkable since, as has been described in chap. 3.1.2, most of the constructions *ad sensum* with CCN in Modern Romance may, if at all, occur with e.g. a relative clause in between the subject and its predicate. From 64 occurrences of *gent* in subject position in the 12th century, 23 predicates are in the plural and 37 in the singular (in 4 cases, the verb was omitted).

- (93) a. Danz Eneas par mer s'en fuit, *grant gent* a pris en son conduit, naigent et singlent a effors; esloingnié se sont bien des pors. (*Frantext*: Anonymous, *Le Roman d'Eneas*, 1160, p. 56)
 ‘And so Aeneas flees over the sea, he took a *big entourage/many men* with him, they navigate and sail with great effort, they have moved very far away from the ports.’
- b. *Une grant gent* poiez oïr Leens en cel chemin venir. (*Frantext*: Hue de Rotelande, *Ipomédon*, 1180, p. 190)
 ‘You can hear a *big crowd* coming there on this way.’
- c. *la gent* l'apelent Montrevel, mes peres n'a meillor chastel. (*Frantext*: Chrétien de Troyes, *Erec*, 1170, p. 2a)
 ‘(the) *people* call_{pl} it Montrevel, my father has no better castle.’

Many, but not all occurrences like (93c) can be categorised not as CCNs but as SOMNs. This example was interpreted as a SOMN since it is not a delimited collection of persons, contiguous in space and/or time, but rather an indefinite mass of people. As in the case of the many ambiguous occurrences of *mobilier*, this is again not a categorisation based on morpho-syntactic criteria, but rather the overall impression of the context. Over the next centuries, while singular occurrences continue to decrease, there is little evolution in the semantics of *gent* or the determiners that *gent* is combined with. There is the slight tendency towards what could be called contiguous loosening, i.e. a decrease in the occurrences where there is a small number of people tied to a central attraction point, e.g. a monarch's retinue, as well as a slight increase in count determiners in combination with *gent*, but these are only subtle tendencies. This impression is quantitatively represented in Figure 8.16 below.

Before turning to the examination of the plural occurrences of *gens*, a short comment on the singular object noun (SON) occurrences in Figure 8.16 is necessary. There are some very marginal cases of *gent* in the 12th, 13th and 14th centuries

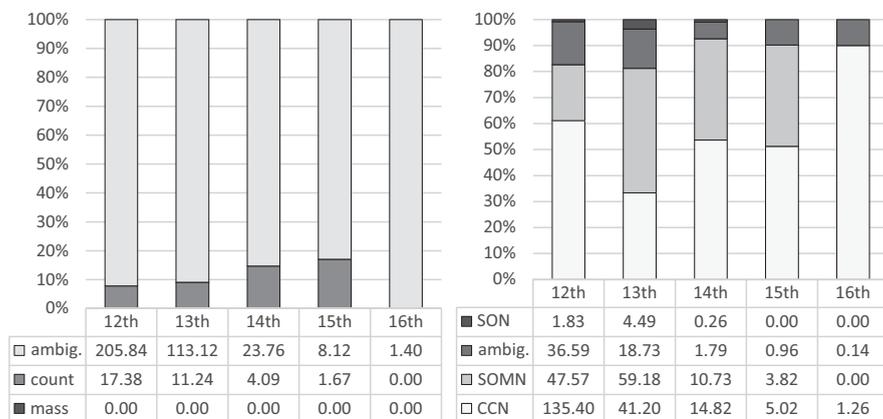


Figure 8.16: Evolution of types of determiner in combination with Fr. *gent*_{SG} and evolution of the nominal types of Fr. *gent*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000).

which can only be interpreted as having singular reference. Two examples may illustrate this:

- (94) a. Car il n'a *gent* an mon ostel An cui ge aie nule atandue [. . .]. (*Frantext*: C. de Troyes, *Yvain ou Le Chevalier au Lion*, 1177, p. 104c)
 'Since there is no *person/nobody* in my house in whom I have no hope [. . .].'
- b. Il descent et vient a la porte; si apele. Quant li portiers entent qu'il a *gent* a la porte, si demande qui il est et que il velt. (*Frantext*: Anonymous, *Le mort le roi Artu*, 1230, p. 56)
 'He goes down and comes to the door, he calls. When the gatekeeper hears that there is a *person/somebody* at the door, he asks who he is and what he wants.'

In both examples of (94), *gent* clearly refers to one single person and may even be interpreted as an indefinite pronoun, *nobody* in the case of (94a) and *somebody* in (94b). Indeed, evident here is the pan-Romance tendency to base e.g. negation particles and indefinite pronouns on either human nouns or body parts. For instance, the It. *niente* 'nothing' and OFr. *néant* 'nothing/not' both go back to Lat. *ne(c) gente(m)* 'no people' (cf. Iliescu 2011; Dessì Schmid 2017) and Catalan has the negative particle *cap* '(lit.) head' (cf. GDC, s.v. *cap*). As indicated in the translations in (94), *gent* may be interpreted as a generic count noun in these contexts; plural reference is thus impossible here. These cases are, however, very marginal (they are 8 occurrences in total) so a conclusive interpretation of them is not possible to achieve.

The plural occurrences of *gens* should be classified as POMNs from the very beginning to present-day French. They always refer to an indefinite yet coherent mass of people who either have a common attraction point – followers – or a common function or task – churchmen, mercenaries. There are only marginal cases of *gens* representing the plural of a CCN up until the 16th century, thus correlating with the disappearance of the inflectional singular having this nominal type. For instance, the prepositional phrase *du monde* ‘of the world’ and the description of a certain kingdom in (95a) below indicate that *gens* in this context does not refer to people in general, but rather to a plurality of other kingdoms or tribes. These occurrences are, however, marginal and there is a steadily low proportion of distributive quantifiers in combination with *gens*, of which most are identifying adjectives like *autre* ‘other’ or *certain* ‘certain’ selecting a particular sub-group from a collection. ‘Real’ distributive quantifiers are mostly *quelques/plusieurs* ‘some_{distr}/various’, but also numerals, then mostly in combination with adjectives like *jeune* ‘young’ or *vieux* ‘old’ and quantifying binominal constructions like *un grand nombre de* ‘a big number of’. Figure 8.17 shows the proportions of ambiguous versus distributive determiners in combination with *gens* over the centuries. These figures include the pseudo-distributive determiners *autre* and *certain*, but not the binominal constructions.

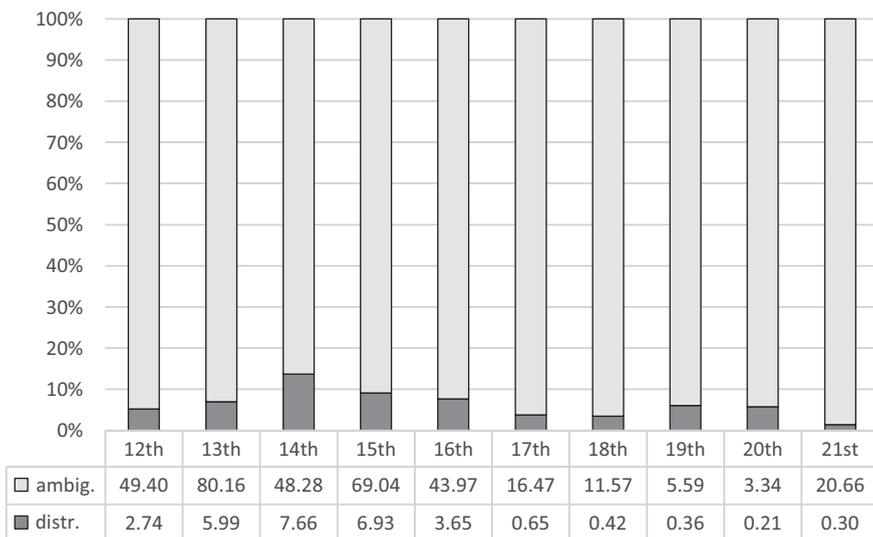


Figure 8.17: Evolution of types of determiner in combination with Fr. *gens_{pl}* in *Frantext* (rel. frequencies extrapolated to 1,000,000).

As in the case of the singular occurrences, *gens* in the plural also undergoes a semantic change towards contiguous loosening. Until the 16th century, *gens* was often determined by a possessive pronoun and/or modified by prepositional phrases indicating the affiliation of them. From the 17th century onwards, there is a tendency for *gens* to occur without a modifier and occurrences with the partitive article increase, pointing to a semantic change towards *gens* referring to an indefinite mass of people. Very common are therefore existential constructions like *il y a des gens qui* ‘there are people who’ (cf. (95b)). Nevertheless, *gens* may always only refer to either (adult) men or mixed groups, but not to women alone (cf. (95c)).

- (95) a. Et eulx mesmes de leur condition se combatent en leur pays les ungs contre les aultres, et se donnent de grans batailles, et est telle la condition de ce dict royaume, et font guerre à tous les *gens* du monde par mer et par terre. (*Frantext*: G. Le Bouvier, *Le Livre de la description des pays*, 1451, p. 120)
 ‘And they themselves, people of this type, fight in their lands one against the other, and they fight great battles, and such is the property of the said kingdom, and they wage war on all the *tribes/nations* of the world on sea and land.’
- b. Quand je pense qu’il y a des *gens* qui sont racistes. (*Frantext*: A. Page, *Tchao pantin*, 1982, p. 170)
 ‘When I think that there are *people* who are racist.’
- c. et s’il n’y a plus assez de jeunes *gens* pour toutes les filles, nous pouvons très bien rester pour compte. (*Frantext*: A. Maurois, *Les Silences du Colonel Bramble*, 1918, p. 124)
 ‘and if there are not enough young *men* for all the girls anymore, we are very well left behind.’

Summarising the analysis of the evolution of *gens* from Old to Modern French, it can be stated that *gent* in Old French fluctuated between use as a CCN and as a SOMN; the singular occurrences decreased, however, until this inflectional singular disappears in the 16th/17th century. Parallel to this development, there is an increasing number of occurrences in the plural which from the noun’s earliest appearances refer to an indefinite mass of people and that may therefore be classed as POMNs. The marginal occurrences of *gens* with distributive quantifiers, especially with small numerals, may point to the beginnings of an evolution of a distributive plural that may in turn recreate an equivalent singular form (cf. Figure 7.2), but this evolution, however, has not become apparent. Of special interest would be in this respect a systematic comparison to diachronic data on

Sp. or It. *gente* which share the same Latin etymon with Fr. *gens*, but which did not fossilise the inflectional plural (cf. Ising 2019, 333–343 for a corpus analysis on the diachronic evolution of It. *gente*, cf. Mihatsch/Kleineberg 2022 for an analysis on the diachronic evolution of Sp. – amongst others – *gente*). Concluding this section of the evolution of *gens*, a short note on a possibly parallel development is necessary. It was indicated regarding example (94) that *gent* in Old French occurred in some very marginal cases that may be interpreted as some pre-stage to an indefinite pronoun or negative particle. These contexts were, however, not continued. For Modern French, Schnedecker (2012; 2015) as well as Cappeau/Schnedecker (2014a) assume that *gens* is on its way to becoming pronominalised, a process which is said to be favoured by oral communication. Indicators for this possible parallel pronominalisation are said to be amongst others non-modified occurrences, occurrences with determiners neutral to the mass-count distinction (e.g. *beaucoup de* ‘much/many’) as well as occurrences with generic reference. The full noun *gens*, in contrast, is said rather to be modified by e.g. adjectives and prepositional phrases, to be more combined with determiners implying countability (with the restrictions imposed by the status as a POMN), and to occur increasingly with specific reference (cf. Cappeau/Schnedecker 2014a, 64–68 for a systematic comparison). In contrast to the pronominalised *a gente* in Brazilian Portuguese, however, these tendencies are indeed only timid indicators and as Schnedecker (2015, 264) puts it: “Du coup, le chemin qui mènerait *gens* à la pronominalisation paraît encore bien long. . .” (‘As a result, the path which would lead *gens* to pronominalisation still seems to be very long’).

For *mobilier* and *gens* discussed in this chapter, the following it can now be stated: although they lack an eventive origin in not being associated with a certain action, both show, to a certain degree, sections of the assumed lexicalisation path of collection nouns. This leads to a confirmation of the hypothesis of Grimm/Levin (2012) that at least typical OMNs are artefact nouns and may thus share a common evolutionary path. It is, however, not clear, how *gens* fits in with this picture. A crucial follow-up analysis should also take into consideration the analysis of Sp. or It. *gente* which have fossilised the singular but not the plural form.

8.3.3 Collection nouns based on the Latin neuter plural: *vaisselle*, *bétail*, *volaille*

The last category of collection nouns is represented by three examples that go back to the Latin neuter plural. As has been described above, *vaisselle* is the only collection noun that represents a direct continuation of a Latin etymon, while

bétail and *volaille* are cases of analogy by means of the pivot of Lat. *animalia* ‘animals’.

Vaisselle: For *vaisselle*, there is not much change detectable over the centuries. The collection noun is first attested in the 12th century, when it appears predominantly in the singular (cf. Figure 8.18).

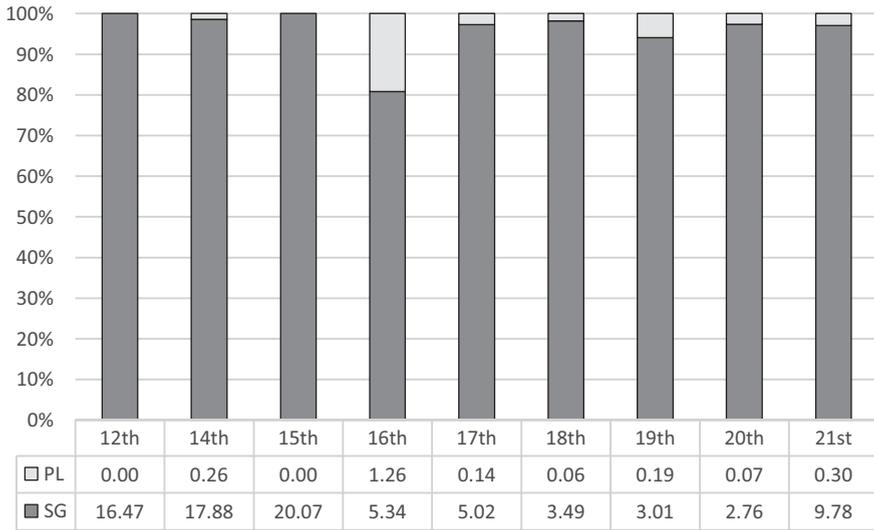


Figure 8.18: Evolution of number inflection of Fr. *vaisselle* in *Frantext* (rel. frequencies extrapolated to 1,000,000).

With the inflectional singular it nearly always refers to an unbounded collection of entities and thus should be categorised mostly as a SOMN. As such, it is combined with the partitive article from the 14th century on (cf. (96a)); the proportion is, however, constantly low. There are some marginal cases of *vaisselle* being combined with count determiners where there is either a sortal reading of the SOMN (cf. (96b)), an interpretation as a singular object noun (cf. 96c)), or as a CCN (cf. (96d)).

- (96) a. Il estoit une fois un homme qui avoit de belles maisons à la ville et à la campagne, de la *vaisselle* d'or et d'argent, des meubles en broderies, et des carosses tout dorez. (*Frantext*: C. Perrault, *Les Contes des fées*, 1697, p. 149)
 ‘Once upon a time, there was a man who had beautiful houses in the city and in the country, golden and silver *cutlery*, embroidered pieces of furniture, and a gold-plated carriage.’

- b. Ils ne purent s'empêcher de me dire qu'un voyageur jouait gros jeu en portant avec lui *une pareille vaisselle*, et particulièrement en Italie, où l'on rencontrait des voleurs à chaque pas. (*Frantext*: A.-R. Lesage, *Histoire de Guzman d'Alfarache*, t. 4, 1732, p. 34)
 'They could not restrain themselves to tell me that a voyager risks much carrying with him *such a kind of cutlery*, particularly in Italy where there are thieves everywhere'
- c. [. . .] apres qu'il m'eut fait apporter une belle grande *vaisselle de terre*, dans laquelle j'arrangeay tout mon cas. (*Frantext*: J. de Léry, *Histoire d'un voyage fait en la terre du Brésil (1578)*, 1580, p. 463)
 '[. . .] after he had a beautiful large *earthen vessel* brought, in which I arranged all my belongings.'
- d. [. . .], à quarante-deux desquels elle a fait présent d'une *vaisselle d'argent* de cent couvert. (*Frantext*: C.-J. de Ligne, *Lettres à la marquise de Coigny pendant l'année 1787*, 1787, p. 59)
 '[. . .], to forty-two of them she gave a *set of silverware* of one hundred pieces as a present.'

There is no tendency of one or the other type to systematically increase or decrease over time and the ambiguous determiners prevail in all occurrences analysed in the corpus.

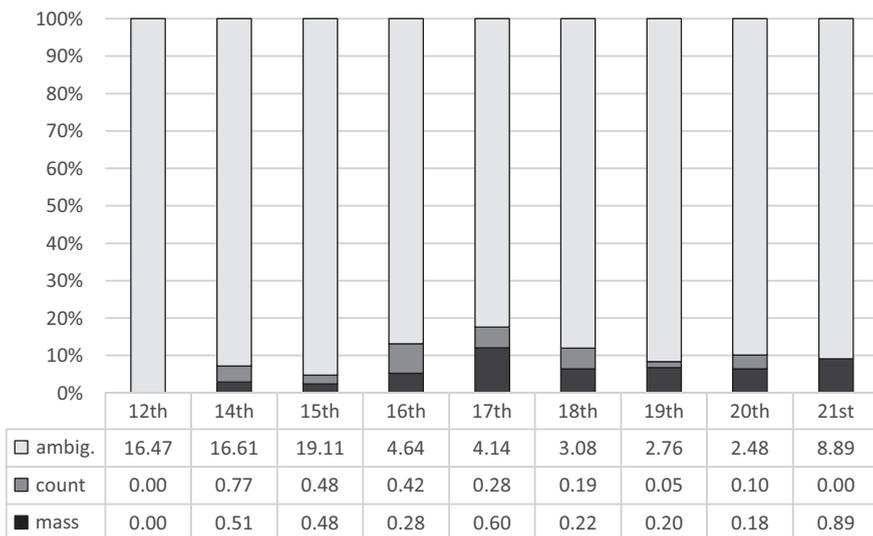


Figure 8.19: Evolution of types of determiner in combination with Fr. *vaisselle*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000).

In its origins, *vaisselle* mostly denotes silverware and gold plates and is thus often modified by prepositional phrases like *d'argent* '(of) silver' and *d'or* 'golden' (cf. also (96a)). Over time, the proportion of unmodified occurrences of *vaisselle* increases and it increasingly denotes tableware or crockery. From the 19th century on, there is a parallel evolution of *vaisselle* first meaning 'dishes' and then washing the dishes, *faire la vaisselle*. The marginal cases of *vaisselle* with plural inflection mostly refer to single recipients or pieces of tableware:

- (97) J'y trouvai [. . .], plusieurs meubles et *vaisselles d'argent* qu'on sauvait chez lui. (*Frantext*: Mme de Sévigné, *Correspondance*, t. 1: 1646–1675, 1675, p. 164)
 'There, I found various pieces of furniture and *silver vessels* which were saved at his home.'

Summarising, Fr. *vaisselle* represents a direct continuation of the Lat. neuter plural *VASCELLA*, and is attested as a SOMN from the very first appearances in the *Frantext* corpus, but there are no evolutionary tendencies with respect to its nominal type or the determiners it is combined with, nor is there any crucial semantic change.

Bétail: The same applies to Fr. *bétail* 'cattle'. *Bétail* is first attested in the *Frantext* corpus in the 14th century. There are some occurrences in the plural in the first two centuries but, as shown in Figure 8.20 below, *bétail* is from the 16th century to nearly the present day attested exclusively in the singular.

The medieval occurrences of *bétail* in the plural refer to single animals (cf. (98a)). There are also two occurrences of *bétails* in the 21st century that should be interpreted as sortal readings (cf. (98b) and (98c)).

- (98) a. il puisse ses diz *bestaulx* envoyer en la dicte pasture, comme il a acoustumé. (*Frantext*: Anonymous, *Le Canarien*, *Pièces justificatives*, 1327, p. 438)
 'he may send his said *heads of cattle* to the said meadow, like he used to.'
- b. l'hiver, ils réparent leurs outils, leurs habitats, et fabriquent des jouets en bois, *bétails*, attelages, carrioles, tombereaux. (*Frantext*: P. Guyotat, *Formation*, 2007, p. 92).
 'during the winter, they repair their tools, their housing and they craft wooden toys, *different kinds of cattle*, trailers, carts, tipcarts.'
- c. [. . .] disséminant la maladie mortelle aux *bétails*, aux volailles. . . (*Frantext*: A.-M. Garat, *L'enfant des ténèbres*, 2008, p. 272)
 '[. . .] distributing the mortal illness to all *kinds of cattle*, of poultry. . .'

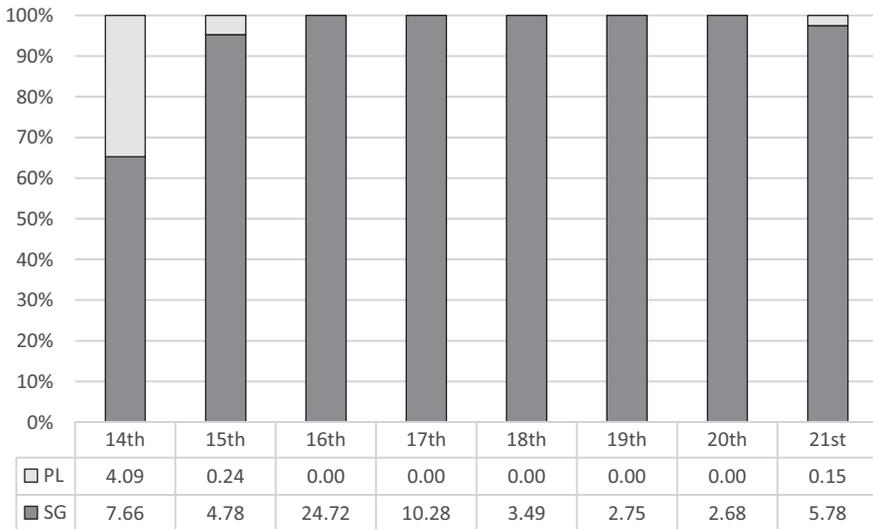


Figure 8.20: Evolution of number inflection of Fr. *bétail* in *Frantext* (rel. frequencies extrapolated to 1,000,000).

The inflectional singular of *bétail* is predominantly used as a SOMN and, as such, there are first attestations of classifier constructions in the 18th century (cf. (99a)) and a small but constant proportion of occurrences with the partitive article marking mass syntax (cf. (99b)). The determiners marking count syntax as represented in Figure 8.21 below are mostly cases of sortal readings (cf. (99c)) and very marginally also of interpretations as a singular object noun as in ‘head of cattle’. Example (99d) is very illustrative in this respect, since there are two occurrences of *bétail* contrasted, first as a singular object noun and then as a SOMN.

- (99) a. J’ai déjà parlé de l’évaluation par tête de bétail. (*Frantext*: A.-R.-J. Turgot, *Réflexions sur la formation et la distribution des richesses*, 1766, p. 146)
 ‘I have already spoken of the evaluation per head of cattle.’
- b. Mais il va me falloir du fumier en masse, et pour avoir du fumier, il faut du bétail. (*Frantext*: É. Erckmann, *Histoire d’un paysan*, t. 1, 1870, p. 382)
 ‘But I will need a large amount of manure, and to have manure, one needs cattle.’
- c. L’éleveur dispose à notre époque, dans un nombre croissant de pays, d’un bétail de valeur et de moyens efficaces pour le défendre contre la maladie. (*Frantext*: M. Wolkowitsch, *L’Élevage dans le monde*, 1966, p. 93)

‘The breeder has, in our days, in an increasing number of countries, *a kind of valuable cattle* and effective means of protecting it against illness.’

- d. Cecy doit estre fait quand il n’y a *aucun bestiail* en ta maison: car ou il y a *du bestiail* on peult faire nettoyer tous les jours la cuisine, [. . .] (*Frantext*: C. Cotereau, *Les douze livres de Lucius Junius Moderatus Columella des choses Rusticques*, 1551, p. 97)

‘This has to be done when there is no *head of livestock* in your house, since, where there is *livestock*, one can clean the kitchen every day, [. . .]’

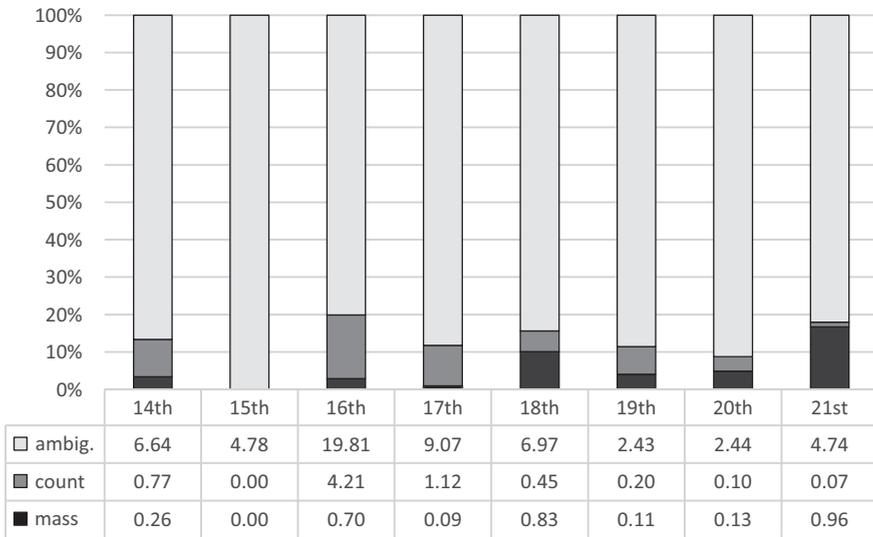


Figure 8.21: Evolution of types of determiner in combination with Fr. *bétail*_{sg} in *Frantext* (rel. frequencies extrapolated to 1,000,000).

There is no semantic change in the evolution of *bétail*: it always refers to cattle or more generally to livestock in contrast to non-domestic animals. Summarising, as in the case with *vaisselle*, there is no apparent evolution in *bétail*, neither with regard to its syntactic nor to its semantic characteristics. As such, it does not follow the assumed lexicalisation path of collection nouns.

Volaille: The last collection noun examined shows a very similar pattern in diachrony. *Volaille* fluctuates between the inflectional singular and plural from its first attestation to the more recent texts in the *Frantext* corpus (cf. Figure 8.22 below).

In contrast to *vaisselle* and *bétail*, there is no predominance for one or the other type in the case of *volaille*. In contrast to its Romance translation equivalents like It. *pollame* or Sp. *volatería*, but also like Engl. *poultry* and Germ. *Geflügel*, Fr. *volaille* is not a typical OMN and never has been.

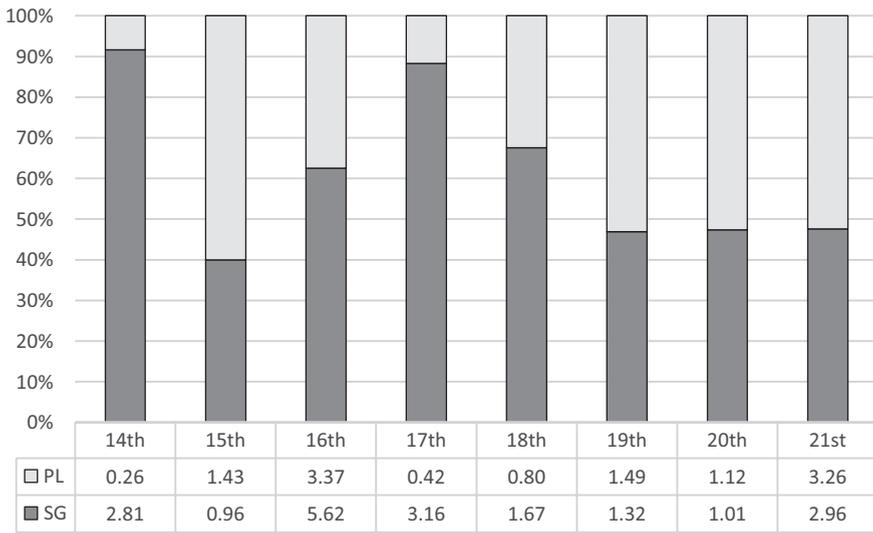


Figure 8.22: Evolution of number inflection of Fr. *volaille* in *Frantext* (rel. frequencies extrapolated to 1,000,000).

The corpus analysis shows that *volaille* is polysemic between usage as a SOMN and an interpretation as a singular object noun over the whole period of time. Consequently, the plural occurrences of *volaille* all refer to single animals (cf. (100a)) or cooked parts of them (cf. (100b)) and the singular occurrences fluctuate between an interpretation as a mass superordinate (cf. (100c)) or as single animals (cf. (100d)).

- (100) a. Et plus au dela un puy, avec deux ou trois grandes auges de pierre de taille pour abreuver le bestial et les *volailles*, [. . .] (*Frantext*: C. Estienne, *L'Agriculture et maison rustique*, 1564, p. 6 v^o)
 'In addition, a well, with two or three big stone troughs to water the cattle and the *poultry*, [. . .]'
- b. Pour moi, je la suivis dans la cuisine, où elle me mit aux mains avec un reste de ragoût de la veille et des *volailles* froide, une bouteille de vin presque pleine, et du pain à discrétion. (*Frantext*: P. de Marivaux, *Le Paysan parvenu*, 1734, p. 49)

‘For me, I followed her into the kitchen, where she competed with me for the rest of yesterday’s ragout and some cold *pieces of chicken*, a nearly full bottle of wine and some bread at will.’

- c. Aussi avoyent une autre manière de *volaille* qu’ils nommoient Poulle Chalcidique, et qui approchoit grandement des meurs des Poules Tanagriques. (*Frantext*: P. Belon, *L’histoire de la nature des oyseaux*, 1555, p. 246)

‘They also had another kind of *poultry* which they named Halkidiki chicken and which was very similar to the Tanagra chicken.’

- d. Ceux qui les remplaçaient dans le train pour rejoindre de proches localités portaient des corbeilles de jonc, des musettes, de vastes paniers d’osier à couvercle d’où émergeait parfois la tête d’une *volaille*. (*Frantext*: R. Sabatier, *Les Noisettes sauvages*, 1974, p. 13)

‘Those who replaced them on the train to nearby towns carried baskets of rushes, musettes [i.e. bagpipes], large wicker baskets with lids from which the head of a *chicken* sometimes emerged.’

The ambiguity illustrated by (100) is also reflected by the determiners which are mostly undetermined with respect to the mass-count distinction, but the inflectional plural is also found with count as well as mass determiners.

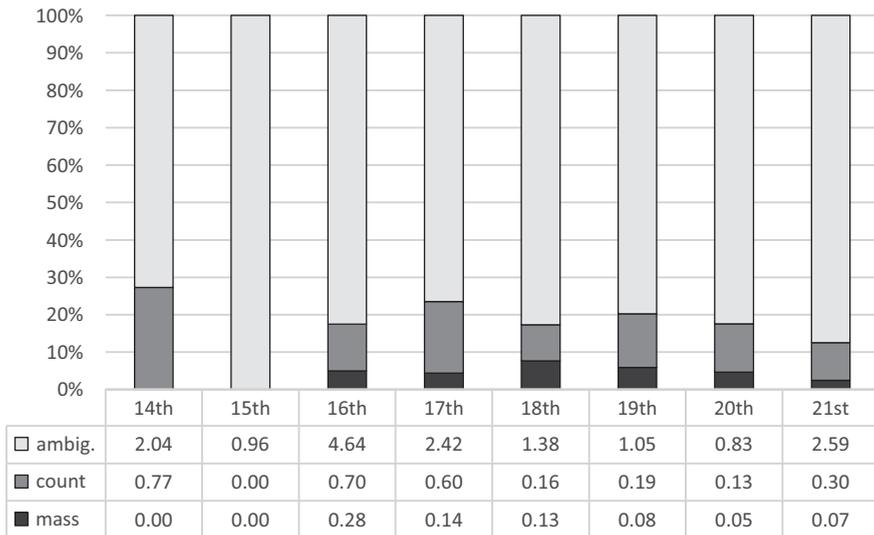


Figure 8.23: Evolution of types of determiner in combination with Fr. *volaille*_{SG} in *Frantext* (rel. frequencies extrapolated to 1,000,000).

Summarising this analysis of these three collection nouns going back to the Latin neuter plural, it can be confidently said that they do not follow the same pattern of evolution as the other collection nouns discussed in chaps. 8.3.1 and 8.3.2. Fr. *bétail* and *vaisselle* are SOMNs over the whole period of time analysed and *volaille* represents a polysemous noun fluctuating between a singular object noun and a SOMN. In neither of the three collection nouns examined was there an apparent evolution in a certain direction, which is to say that the nouns are of the same aspectual type from their first attestations in the corpus up until their use in Modern French.

9 Conclusion and discussion: Linguistic characterisation of collection nouns in language evolution

The last part of this study examined the diachronic development of French collection nouns. Research done until now on this subject has concentrated mainly on the origins of collection nouns. It was stated in this respect that especially OMNs are typically artefact nouns originating in an eventive context as participants in a certain kind of action. Following research on historical linguistics and in particular on the work of Mihatsch (2016), it was then assumed that collection nouns typically come into being as deverbal derivatives meaning first the action itself and then the artefacts involved in that action. These artefacts then get more and more coherent until they become a bounded collection, verbalised by a CCN. Over time, there is the assumption of the referents becoming increasingly homogeneous and the outer boundaries more and more permeable until a SOMN evolves. The more homogeneous the referents become, the more the degree of distributivity of the OMN increases until there may be a POMN and even the recreation of an inflectional plural, now as an object noun. Particularly the first stages of evolution are attested for a number of collection nouns in various languages analysed especially in the context of traditional historical linguistics. Research on the acquisition of nouns in general and of collection nouns in particular additionally supports this theory. The main research question for this diachronic part consequently addressed the issue of finding empirical evidence to buttress the theory of a uniform lexicalisation path of collection nouns. It was furthermore assumed that non-artefactual nouns, whose referents typically do not participate in actions, may – if at all – pattern with artefactual nouns only by analogy (cf. HP_{dia}1). Collection nouns going back to the Latin neuter plural were moreover assumed not to pattern with derived nouns since they also lack the possible relation to an event (cf. HP_{dia}2).

With regard to the main research question, the lexicalisation path of derived artefactual collection nouns could be confirmed – but only to a certain degree. Fr. *habillement* ‘clothing’ and *accoutrement* ‘attire’ matched the developmental stages assumed for them. Fr. *équipement* ‘equipment’ and *vêtement* ‘clothing/piece of clothing’ were, however, said to either be already ahead, before or somewhere next to the assumed pathway. For Fr. *mercerie* ‘haberdashery’ and *fringues* ‘clothes’ there was unfortunately not enough data available to formulate any consistent conclusion. It may thus be stated that the assumed pathway of lexicalisation is empirically confirmed, though with the reservation that the assumed stages

very often appear parallel to one another, that there are a lot of ambiguous cases clearly not analysable and most importantly, that the assumed pathway could only be examined on the basis of two collection nouns. Together with the state of the art, I will, however, suppose that it may also be true for other cases that are similar to *habillement* and *accoutrement*. This is also supported by the analysis of collective nonce-formations in chap. 5. On the one hand, the collective nonce-formations found in the *TenTen* corpora of French, Spanish and Italian have predominantly inanimate, artefactual referents, underlining the strong link between the concepts of ACTION and COLLECTION – even with functional collection nouns. On the other, there were no clear cases of mass superordinate interpretations in any of the three corpora. It was argued that is mainly due to pragmatics and the contextual bounding of the referents. Together with the diachronic analysis, it can now be stated that the missing OMN category in the morphological analysis may also be ascribed to the fact that OMNs only arise through diachronic evolution.

With respect to the ontological type of referents as well as the morphological form of a collection, the hypotheses formulated above cannot be answered with certainty. The analysis of *bétail* ‘cattle/livestock’ and *volaille* ‘poultry’ showed that these two collection nouns do not match the pattern of the assumed lexicalisation path, as there was no apparent evolution detectable. It has been assumed that animate collection nouns only take part in the assumed lexicalisation path – if at all – by analogy. This would mean that the reason for the lack of matching developments in the case of *bétail* and *volaille* lies in their referring to animates. However, at the same time they are not typical derivatives, i.e. those formed by means of a certain suffix, but continuations of their respective Latin etymon in its neuter plural form. Strictly speaking, the analysis thus cannot answer the question of whether their deviance is caused by the ontological type of referents or rather by their word-formation pattern. Given the fact that *vaisselle* does not follow the assumed lexicalisation path either, I formulate the cautious follow-up hypothesis that it is not the ontological type that determines the diachronic evolution of a collection noun, but more importantly its etymological origin.

IV General conclusion and outlook

10 Summary of the results and conclusion

This present study started from the basic insight that extra-linguistic entities may be construed in various ways by using different linguistic means of expression. This is valid not just for objects but also for abstract concepts and actions. The linguistic construal of extra-linguistic reality is best described with the notion of nominal aspectuality. This concept was fleshed out herein on the basis of various research approaches from cognitive linguistics and typology, as the linguistic modelling of external boundedness and the internal structure of entities. Parallel to an onomasiological approach to verbal aspectuality, nominal aspectuality is primarily understood as a conceptual model in which the construal of extra-linguistic entities may be verbalised by various means of linguistic expression. In this sense, the conceptual external boundedness of an entity correlates with mostly grammatical means like the possibility of an inflectional plural, but also with lexical means of expressing aspectuality such as the fact that certain adjectives may only predicate a whole entity and not its constituting parts. While it seems to be a cross-linguistically valid factor that individual entities are denoted by count nouns and substances by mass nouns, it is plural entities in particular that are of linguistic interest. Languages differ especially in their grammatical means of expressing external boundedness, i.e. in their nominal systems. While most languages of the world distinguish in some way between singular and plural, and in the demarcation of count and mass nouns, they vary in their actual means of expression. These means may range from classifier constructions to plural marking by way of reduplication. Even in the family of Romance languages, there are clear differences in the distinction between mass and count nouns. By taking into consideration the four major Romance languages of French, Italian, Spanish and Portuguese, the description of their particular nominal systems led to a Romance continuum of the marking of the mass-count distinction going from French, which overtly marks mass vs. count nouns, then to Italian, Spanish European and finally to Brazilian Portuguese, which allows for this distinction to be neutralised – nouns may appear as bare singulars referring to a plurality of entities. Whereas the Romance languages examined here differ in their linguistic expression of boundedness, their means of expressing an entity's internal structure were argued to be cross-linguistically more or less equivalent. For instance, all the languages examined distinguish between nouns denoting entities without an internal structure (*boy*) and those that have one (*team*).

Adopting this conception of nominal aspectuality, the aspectual type of collection nouns could then be defined. Collection nouns are thus nouns that in their morphologically default form refer to a plurality of entities. In contrast to the

distributive plural expressed by the inflectional plural, collections denoted by collection nouns cohere to a greater degree. Collection nouns may differ in their morphological form (singular with an equivalent plural form or transnumeral singular or plural only) and their syntactic-semantic characteristics, but they all have in common the feature of internal plurality. In contrast to aggregates, they display some sort of internal hierarchical structure. Collection nouns are an intriguing field of research since they display a mismatch between the internal plurality of entities and their outer coding as either singular or transnumeral.

With respect to Romance collection nouns, various research gaps were identified. These gaps related to the synchronic characteristics of present-day language as well as to the diachronic evolution of collection nouns. With respect to collection nouns in present-day language, there has been hardly any research on Romance OMNs and no comparative studies taking into consideration more than one type of collection noun and more than one (Romance) language. Furthermore, Romance collection nouns have generally not been considered from a derivational point of view since they are said not to represent a particularly productive word-formation pattern. With respect to the diachronic evolution of collection nouns, there are at present some studies examining mainly their origins, but until now there have been no exhaustive empirical studies focusing on the different stages of development of the various types of collection noun. On the basis of these three research gaps, the present thesis addressed three major research questions:

RQ1: What are the influencing factors on the particular linguistic expression of a collection of entities and the semantic-syntactic characteristics related to it?

RQ2: To what extent are there any productive word-formation patterns in the domain of COLLECTIVITY?

RQ3: To what extent do collection nouns follow a unidirectional path of lexicalisation? Can this path be empirically proven?

To address the first main research question (RQ1) and to thus examine the semantic-syntactic means of expressing COLLECTIVITY in Romance languages, a series of acceptability judgement studies was conducted. Specifically, the studies looked at the question of whether there are systematic differences between various aspectual nominal types, whether are cross-linguistic differences exist in the characteristics of OMNs in particular, as well as whether the ontological type of its constituting entities influences the semantic-syntactic features of a collection noun. The studies confirmed systematic differences between various collection noun types depending on their constitution of nominal aspectuality. Typical CCNs denote bounded col-

lections whose constituting entities are structured by meronymic links. Because of their conceptual boundedness, they are countable, but their internal plurality is not easily accessible. SOMNs denote unbounded collections whose constituting entities may be structured both by meronymic and hyponymic relations. This unboundedness leads to an uncountability of the collection nouns, but also to a compatibility with certain lexical means of expressing the distributive plural like stubbornly distributive predicates and to a lesser extent other highly distributive predicates such as ‘one after the other’. Finally, POMNs also denote unbounded collections but in contrast to morphologically singular mass collections, the fossilised inflectional plural facilitates to a greater extent linguistic access to their internal plurality. The POMNs examined in these present studies are mostly characterised by a hyponymic relation between the collection itself and its constituting members. Cross-linguistic variation depending on their differences in the marking of the mass-count distinction could be confirmed, but only for the semantic domain of inanimate artefact nouns. The ontological type of the constituting entities does not seem to play a role for the semantic-syntactic properties of collection nouns in Romance languages. The results of the studies taken together led to a revised schematic model of collection nouns in the Romance languages, justifying on the one hand their subsumption under the umbrella term of collection nouns, but on the other hand, also displaying the systematic differences between them (cf. Figure 4.15).

The second research question (RQ2) concerned Romance collection nouns as a derivational category. The state of the art suggested in this respect that Romance collection nouns do not represent a very productive word formation pattern. Research done so far, however, only focused on the past, i.e. the realised productivity of certain collectivising suffixes, by considering mostly the number of types listed in dictionaries. The present study examined in addition their expanding productivity by adopting the theoretical basis of construction morphology and by analysing nonce-formations formed by means of French, Spanish and Italian suffixes. The results showed that the collective suffixes in French do not show any previous nor currently expanding productivity. The concept of COLLECTIVITY is expressed rather by analytic means of expression in this Romance language, for instance through of binominal constructions like *groupe de N* ‘group of N’. In contrast, Spanish makes use of the suffix *-erío* which not only in Mexican Spanish but also in any other diatopic variety forms new collection nouns to a considerable extent. Italian in addition has three collectivising suffixes, *-ame*, *-ume* and *-aglia*, where all three are able to form new collection nouns. These productive suffixes mostly directly collectivise the referents of the derivational base noun. This was the case with Sp. *tumberío* ‘graveyard’ in (61b), It. *gazzaglia* ‘swarm of magpies’ in (67c) and of It. *oggettume* ‘collection of things’ in (68a). In this respect, they are typical representatives of what Baldinger (1950) calls functional collectivity.

More marginally, they may also trigger an associative plural reading by collectivising the referents associated with a central focal point. This was the case with It. *mitrokinaglia* ‘everything that has to do with Vassili Nikitich Mitrokhin’. This function was, however, more associated with other, only secondarily collectivising suffixes as in Sp. *Fujimontesinada* ‘people associated with Alberto Fujimori and Vladimiro Montesinos’ in (60b) or It. *cuggineria* ‘people associated with the fashion brand Gucci’ in (65c). The corpus study thus showed that one has to assume a constructional schema <collection> also in Romance languages that has different instantiations and which is productive to varying degrees. All possible collectivising functions diachronically show a systematic network governed by metonymic shifts and analogy which was depicted in this thesis by means of a semantic map of COLLECTIVITY (cf. Figure 5.10).

With respect to the last research question (RQ3), a diachronic corpus study concentrating on French aimed to empirically examine the assumed unidirectional pathway of collection nouns. The analysis of a number of collection nouns with respect to their etymology as well as their diachronic evolution from the very first attestation until the present revealed that the theory of a lexicalisation path of collection nouns is empirically proven only to a certain extent. Only the derived artefact nouns examined showed the assumed stages of an action, functional, count collective noun to becoming, in some cases, an OMN. For all the nouns studied in this part, there were, however, always parallel evolutions and branching strands of development. In each case, the corpus analysis also showed that there were instances of the collection noun in question as object nouns, in both the singular and the plural, which had not been considered in terms of the presumed pathway of lexicalisation. The two cases of Fr. *mobilier* and *gens* equally displayed the change from a CCN to a SOMN and further to a POMN in the latter case. The three collection nouns that were derived from the Latin neuter plural did not show any of the assumed stages. They are either already determined for some nominal aspectual type from their recorded beginnings, as was the case with *bétail* and *vaisselle*, or they flit, like *volaille*, between singular and plural occurrences and accommodate various possible interpretations over the whole time period examined. The analysis thus confirmed the assumption of Grimm/Levin (2012) that typical OMNs are artefact nouns. These show at least to some extent the assumed stages of development, in originating as objects which take part in an action and which then lose more and more of their spatio-temporal bounding. The analysis, however, also showed that a certain stage of this development does not necessarily disappear after the new stage has been attested. Most of the nouns examined show parallel evolutions of more than one nominal aspectual type resulting in a layering of semantically connected meanings, and are thus polysemic nouns. The results therefore suggest that especially OMNs are

a linguistic compromise which may only come into being by lengthy diachronic processes of language change. This is also supported by the corpus analysis of collective nonce-formations where only spatio-temporally bounded collection nouns were found in the corpora.

The present study thus showed that the expression of COLLECTIVITY in French, but also in the other Romance languages investigated here, is a highly diverse phenomenon which is, at the same time systematically governed by very general principles of cognition. By using different kinds of collection noun, we may construe extra-linguistic collections in different ways. The acceptability judgement studies conducted showed that it is essential not only to consider one type of collection noun alone, but to always analyse them as particular types on a nominal aspectual continuum. The results equally proved the importance of considering not only a single language in isolation, since diverse language-specific nominal systems greatly influence the linguistic properties of various types of collection noun. I was furthermore able to show that collection nouns are a productive word-formation pattern, at least in Spanish and Italian. Finally, the diachronic corpus analysis led to a refining of the assumed pathway of evolution of collection nouns. The study consequently gave a fuller picture of how COLLECTIVITY has been expressed over time in French in comparison to Spanish, Italian and Portuguese. It could be shown that there is indeed a continuity between different types of collection nouns, continuity between different Romance languages and continuity in historical development.

11 Outlook

For future research various analyses were outlined in the respective chapters. With regard to the semantic-syntactic characteristics of various collection noun types, several inconsistencies that were still present even in the follow-up study should be addressed. On the one hand, the study overall adopted a very symmetric test design sticking to the three main semantic domains to systematically examine the different linguistic means of expressing the same underlying concept. This strategy was justified in designing the study, but it also led to issues relating to the connotations of the nouns tested and their frequencies in use. Some nouns were thus rated lower than expected since they did not exactly match the context connotationally. Future research should thus refine the acceptability judgement tests in focussing more on connotationally and not necessarily denotationally equivalent nouns. On the other hand, a consideration of Romance-based creole languages would also be fruitful given their specific nominal systems and often only optional marking of nominal number. In theory, they should reflect even more markedly the tendencies already shown by Brazilian Portuguese. With respect to the analysing collection nouns as a derivational category, future research should refine the evaluation of the nonce-formations found in the *TenTen* corpora. It was not, for instance, possible to systematically examine the extent to which the different collective suffixes and their derivational potential differ in their diatopic but also diastratic distribution. This meant that the only method for examining diatopic variation was to consult the domain names for the various corpora. It was furthermore assumed that especially for Italian, teenager's language is particularly prone to these kinds of nonce-formations; however, it was not possible to analyse in detail whether this also applied to the nonce-formations found in the corpora. In addition, a detailed analysis of the aspectual type of derived nouns found could not be undertaken. Future research could thus present the nonce-formations found in the corpus analysis to native speakers of different dialects and across all age groups to ask respondents for their interpretations as to whether they find the nonce-formations transparent, whether they think the nouns denote specifically delimited groupings of individual entities or rather indefinite masses and so on. Finally, further research should broaden the diachronic corpus analysis regarding the evolution of collection nouns, since the present analysis was obliged to focus on only a small number of French collection nouns, neglecting other Romance languages like Italian or Spanish.

12 References

12.1 Corpora

- CDE = Davies, Mark (ed.), *Corpus del español*, 2001–, <https://www.corpusdelespanol.org/> [last access: Mar. 25th, 2022].
- CDP = Davies, Mark (ed.), *O corpus do português*, 2006–, <https://www.corpusdoportugues.org/> [last access: Mar. 25th, 2022].
- CEFC = ORFÉo (ed.), *Corpus d'étude pour le français contemporain*, 2001–, <https://repository.ortolang.fr/api/content/cefc-orfeo/11/documentation/site-orfeo/index.html> [last access: Mar. 25th, 2022].
- CLIPS = Albano Leoni, Federico (ed.), *Corpora e lessici dell'italiano parlato e scritto*, 1999–2004, <http://www.clips.unina.it/it/index.jsp> [last access: Mar. 25th, 2022].
- CDH = Real Academia Española (ed.), *Corpus del diccionario histórico de la lengua española*, 2013, <https://apps.rae.es/CNDHE/view/inicioExterno.view> [last access: Mar. 25th, 2022].
- C-Oral-Rom = Cresti, Emanuela/Moneglia, Massimo (edd.), *C-Oral-Rom. Integrated reference corpora for spoken Romance Languages*, Amsterdam/Philadelphia, John Benjamins, 2005.
- CORIS/CODIS = Rossini Favretti, Rema (ed.), *Corpus di riferimento dell'italiano scritto/Corpus dinamico dell'italiano scritto*, 2001–, http://corpora.dslo.unibo.it/coris_ita.html [last access: Mar. 25th, 2022].
- CORPES XXI = Real Academia Española (ed.), *Corpus del español del siglo XXI. Versión 0.93*, 2021, <https://apps2.rae.es/CORPES> [last access: Mar. 25th, 2022].
- CREA = Real Academia Española (ed.), *Corpus de referencia del español actual. Versión 3.2*, 2008, corpus.rae.es/creanet.html [last access: Mar. 25th, 2022].
- CRPC = Centro de Linguística da Universidade de Lisboa (ed.), *Corpus de referência do português contemporâneo*, 2008–2016, <http://gamma.clul.ul.pt/CQPweb/> [last access: Mar. 25th, 2022].
- Frantext = Analyse et Traitement Informatique de la Langue Française–Centre national de la recherche scientifique (ATILF–CNRS)/Université de Lorraine (edd.), *Base textuelle Frantext*, 1998–, <https://www.frantext.fr> [last access: Mar. 25th, 2022].
- TenTen = *TenTen corpus family*, www.sketchengine.eu [last access: Mar. 25th, 2022].

12.2 Dictionaries and encyclopaedias

- DECH = Corominas, Joan (ed.), *Diccionario crítico etimológico castellano e hispánico. Edition on CD-ROM*, Madrid, Gredos.
- DES = Pharies, David (ed.), *Diccionario etimológico de los sufijos españoles*, Madrid, Gredos, 2002.
- DGL = Pons, Jordi Induráin (ed.), *Diccionario general de la lengua española. El vocabulario general del español de América y España*, Barcelona, Larousse, ²2009.
- DLP = Crystal, David (ed.), *A dictionary of linguistics and phonetics*, Boston/Malden/Oxford, Credo Reference/Blackwell, ⁶2008.
- DPD = Real Academia Española (ed.), *Diccionario panhispánico de dudas*, Madrid, Santillana, 2005.

- DRAE = Real Academia Española (ed.), *Diccionario de la lengua española*, Madrid, Espasa, ²³2014.
- FEW = Wartburg, Walther von, *Französisches Etymologisches Wörterbuch. Eine darstellung des galloromanischen sprachschatzes*, 25 vol., Basel et al., Zbinden et al., 1922–2002, <https://lecteur-few.atilf.fr/> [last access: Mar. 25th, 2022].
- GDC = Fundació Enciclopèdia Catalana (ed.), *Gran diccionari de la llengua catalana*, Barcelona, Enciclopèdia Catalana, 1998, www.enciclopedia.cat/gran-diccionari-de-la-llengua-catalana [last access: Mar. 25th, 2022].
- GDW = Grimm Jacob/Grimm, Wilhelm, *Deutsches Wörterbuch*, 16 vol., Leipzig, Hirzel, 1854–1961.
- GR = Robert, Paul/Rey, Alain, *Le grand Robert. Version numérique*, Paris, Le Robert, 2017, grandrobert.lerobert.com/robert.asp [last access: Mar. 25th, 2022].
- LB = Sauermost, Rolf, *Lexikon der Biologie*, Heidelberg, Spektrum, 1999, <https://www.spektrum.de/lexikon/biologie/> [last access: Mar. 25th, 2022].
- LEW = Walde, Alois, *Lateinisch etymologisches Wörterbuch*, Heidelberg, Winter, ²1910.
- MLS = Glück, Helmut/Rödel, Michael, *Metzler Lexikon Sprache*, Stuttgart, Metzler, 2016.
- MM = María Moliner, *Diccionario de uso del español*, Madrid, Gredos, 1994.
- NDLP = Ferreira, Aurélio Buarque de Holanda, *Novo dicionário da língua portuguesa*, Rio de Janeiro, Nova Fronteira, ²1999.
- OED = Simpson, John/Weiner, Edmund, *Oxford English Dictionary – OED online*, Oxford, Oxford University Press, ³2000, www.oed.com [last access: Mar. 25th, 2022].
- TLFi = Analyse et Traitement Informatique de la Langue Française – Centre national de la recherche scientifique/Université de Lorraine, *Trésor de la langue française informatisé*, <http://atilf.atilf.fr> [last access: Mar. 25th, 2022].
- TLL = Hillen, Michael, *Thesaurus Linguae Latinae. Online edition*, München, Bayerische Akademie der Wissenschaften, 2019, <https://thesaurus.badw.de/tll-digital/tll-open-access.html> [last access: Mar. 25th, 2022].
- Zingarelli = Zingarelli, Nicola, *Lo Zingarelli*, Bologna, Zanichelli, ¹²2007.

12.3 Research literature

- Acquaviva, Paolo, *Lexical plurals. A morphosemantic approach*, Oxford/New York, Oxford University Press, 2008.
- Acquaviva, Paolo, *Singularives*, in: Müller, Peter, et al. (edd.), *Word-formation. An international handbook of the languages of Europe*, vol. 2, Berlin/Boston, De Gruyter, 2015, 1171–1183.
- Aissen, Judith, *Differential object marking. Iconicity vs. economy*, *Natural language & linguistic theory* 21:3 (2003), 435–483.
- Aitchison, Jean, *Words in the mind. An introduction to the mental lexicon*, Malden/Oxford/Chichester, Wiley-Blackwell, ⁴2012.
- Aliquot-Suengas, Sophie, *Référence collective/sens collectif. La notion de collectif à travers les noms suffixés du lexique français*, PhD Thesis, Université de Lille III, 1996.
- Aliquot-Suengas, Sophie, *Les dérivés français à référence collective*, *Langages* 37:152 (2003), 33–50, doi: 10.3406/lgge.2003.2438.
- Allan, Keith, *Collectivizing*, *Archivum linguisticum* 8:2 (1976), 99–117.
- Allan, Keith, *Nouns and countability*, *Language* 56:3 (1980), 541–567.

- Andreou, Marios/Lieber, Rochelle, *Aspectual and quantificational properties of deverbal conversion and “-ing” nominalizations. The power of context*, English language and linguistics 24:2 (2020), 333–363, doi: 10.1017/S1360674319000108.
- Baayen, Harald, *On frequency, transparency, and productivity*, in: Booij, Geert/Marle, Jaap van (edd.), *Yearbook of morphology 1992*, Dordrecht, Kluwer, 1993, 181–208.
- Baayen, Harald, *Corpus linguistics in morphology. Morphological productivity*, in: Lüdeling, Anke/Kyto, Merja (edd.), *Corpus linguistics. An international handbook*, vol. 2, Berlin/New York, De Gruyter, 2009, 899–919.
- Baayen, Harald/Lieber, Rochelle, *Productivity and English derivation. A corpus-based study*, Linguistics 29:5 (1991), 801–843.
- Bach, Emmon, *The algebra of events*, Linguistics and philosophy 9:1 (1986), 5–16.
- Baldinger, Kurt, *Kollektivsuffixe und Kollektivbegriff. Ein Beitrag zur Bedeutungslehre im Französischen mit Berücksichtigung der Mundarten*, Berlin, Akademie, 1950.
- Bale, Alan/Barner, David, *The interpretation of functional heads. Using comparatives to explore the mass/count distinction*, Journal of semantics 26:3 (2009), 217–252.
- Bale, Alan/Gillon, Brendan, *Re-examining the mass-count distinction*, in: Moltmann, Friederike (ed.), *Mass and count linguistics, philosophy, and cognitive science*, Amsterdam/Philadelphia, John Benjamins, 2020, 13–36.
- Barker, Chris, *Group terms in English. Representing groups as atoms*, Journal of semantics 9:1 (1992), 69–93.
- Barner, David/Snedeker, Jesse, *Quantity judgements and individuation. Evidence that mass nouns count*, Cognition 97:1 (2005), 41–66.
- Barsalou, Lawrence, *Ad hoc categories*, Memory and cognition 11:3 (1983), 211–227, doi: 10.3758/BF03196968.
- Bassano, Dominique, *Production naturelle précoce et acquisition du langage. L'exemple du développement des noms*, Revue de linguistique et de didactique des langues 31 (2005), 61–84, doi: 10.4000/lidil.136.
- Bauer, Laurie, *English word-formation*, Cambridge/New York/Melbourne, Cambridge University Press, 1983.
- Bauer, Laurie, *Introducing linguistic morphology*, Edinburgh, Edinburgh University Press, 1992.
- Bauer, Laurie, *Morphological productivity*, Cambridge/New York/Port Melbourne/Madrid/Cape Town, Cambridge University Press, 2001.
- Berko, Jean, *The child's learning of English morphology*, WORD 14:2–3 (1958), 150–177, doi: 10.1080/00437956.1958.11659661.
- Bertinetto, Pier Marco, *Tempo, aspetto e azione nel verbo italiano. Il sistema dell'indicativo*, Firenze, Presso L'Accademia della Crusca, 1986.
- Bertinetto, Pier Marco/Squartini, Mario, *Tense and aspect*, in: Ledgeway, Adam/Maiden, Martin (edd.), *The Oxford guide to the Romance languages*, Oxford, Oxford University Press, 2016, 939–953.
- Beviláqua, Kayron/Pires de Oliveira, Roberta, *Brazilian bare phrases and referentiality. Evidences from an experiment*, Revista letras 90 (2014), 253–275, doi: 10.5380/rel.v90i2.37234.
- Biermann, Anna, *Die grammatische Kategorie Numerus*, in: Seiler, Hansjakob/Lehmann, Christian (edd.), *Apprehension. Das sprachliche Erfassen von Gegenständen*, Tübingen, Narr, 1982, 229–243.
- Blank, Andreas, *Prinzipien des lexikalischen Bedeutungswandels am Beispiel der romanischen Sprachen*, Tübingen, Niemeyer, 1997.

- Blank, Andreas, *Der “Kopf” in der Romania und anderswo. Ein metaphorisches (und metonymisches) Expansions- und Attraktionszentrum*, in: Gil, Alberto/Schmitt, Christian (edd.), *Kognitive und kommunikative Dimensionen der Metaphorik in den romanischen Sprachen*, Bonn, Romanistischer Verlag, 1998, 11–32.
- Blank, Andreas, *Pathways of lexicalization*, in: Haspelmath, Martin, et al. (edd.), *Language typology and language universals. An international handbook*, Berlin/New York, De Gruyter, 2001, 1596–1608.
- Blank, Andreas/Koch, Peter, *Kognitive romanische Onomasiologie und Semasiologie*, in: Blank, Andreas/Koch, Peter (edd.), *Kognitive romanische Onomasiologie und Semasiologie*, Tübingen, Niemeyer, 2003, 1–15 (= 2003a).
- Blank, Andreas/Koch, Peter (edd.), *Kognitive romanische Onomasiologie und Semasiologie*, Tübingen, Niemeyer, 2003 (= 2003b).
- Bloom, Paul, *How children learn the meaning of words*, Cambridge, Massachusetts/London, MIT Press, 2000.
- Bloom, Paul/Kelemen, Deborah, *Syntactic cues in the acquisition of collective nouns*, *Cognition* 56:1 (1995), 1–30.
- Bloomfield, Leonard, *Language*, London, Allen & Unwin, 1933.
- Boas, Hans/Höder, Steffen (edd.), *Constructions in contact. Constructional perspectives on contact phenomena in Germanic languages*, Amsterdam/Philadelphia, John Benjamins, 2018.
- Bollée, Annegret, *Reunion Creole*, in: Michaelis, Susanne Maria, et al. (edd.), *The survey of pidgin and creole languages*, vol. 2, Oxford, Oxford University Press, 2013, 241–249.
- Boogaart, Ronny, *Aspect and Aktionsart*, in: Booij, Geert, et al. (edd.), *Morphology. An international handbook on inflection and word-formation*, vol. 2, Berlin/New York, De Gruyter, 2004, 1165–1180.
- Boogaart, Ronny/Janssen, Theo, *Tense and aspect*, in: Geeraerts, Dirk/Cuyckens, Hubert (edd.), *The Oxford handbook of cognitive linguistics*, Oxford/New York, Oxford University Press, 2010, 803–828.
- Booij, Geert, *Form and meaning in morphology. The case of Dutch “agent nouns”*, *Linguistics* 24:3 (1986), 503–517.
- Booij, Geert, *Construction morphology*, Oxford/New York, Oxford University Press, 2010.
- Booij, Geert, *Morphology in construction grammar*, in: Hoffmann, Thomas/Trousdale, Graeme (edd.), *The Oxford handbook of construction grammar*, Oxford/New York, Oxford University Press, 2013, 255–273.
- Booij, Geert, *Word-formation in construction grammar*, in: Müller, Peter, et al. (edd.), *Word-formation. An international handbook of the languages of Europe*, vol. 1, Berlin/Boston, De Gruyter, 2015, 188–202.
- Booij, Geert, *Construction morphology*, in: Hippius, Andrew/Stump, Gregory (edd.), *The Cambridge handbook of morphology*, Cambridge, Cambridge University Press, 2016, 424–448.
- Borer, Hagit, *In name only*, Oxford, Oxford University Press, 2005.
- Borillo, Andrée, *Statut et mode d’interprétation des noms collectifs*, in: Guimier, Claude (ed.), *Co-texte et calcul du sens. Actes de la table ronde tenue à Caen les 2 et 3 février 1996*, Caen, Universidad de Caen, 1997, 105–121.
- Bosque, Ignacio, *El nombre común*, in: Bosque, Ignacio/Demonte, Violeta (edd.), *Gramática descriptiva de la lengua española*, vol. 1, Madrid, Espasa, 1999, 3–76.

- Bosque, Ignacio, *Reflexiones sobre el plural y la pluralidad. Aspectos léxicos y sintácticos*, in: Casas Gómez, Miguel/Torres Sánchez, María Ángeles (edd.), *Actas de las V Jornadas de Lingüística. Cádiz, 23 y 24 de noviembre de 1999*, Cádiz, Universidad de Cádiz, 2000, 5–37.
- Braga, João Vinicius de Almeida, et al., *Bare singular and bare mass nouns in Brazilian Portuguese. First results of an empirical survey*, *Journal of Portuguese linguistics* 9:1 (2010), 75–94, doi: 10.5334/jpl.111.
- Braun, Friederike/Haig, Geoffrey, *When are German “girls” feminine? How the semantics of age influences the grammar of gender agreement*, in: Bieswanger, Markus/Motschenbacher, Heiko/Mühleisen, Susanne (edd.), *Language in its socio-cultural context. New explorations in gendered, global and media uses*, Frankfurt am Main, Peter Lang, 2010, 69–83.
- Brekle, Herbert/Kastovsky, Dieter, *Wortbildungsforschung. Entwicklung und Positionen*, in: Brekle, Herbert/Kastovsky, Dieter (edd.), *Perspektiven der Wortbildungsforschung. Beiträge zum Wuppertaler Wortbildungskolloquium vom 9.–10. Juli 1976; anlässlich des 70. Geburtstags von Hans Marchand am 1. Oktober 1977*, Bonn, Bouvier, 1977, 7–19.
- Brinton, Laurel, *The mass/count distinction and Aktionsart. The grammar of iteratives and habituals*, *Belgian journal of linguistics* 6 (1991), 47–69.
- Brinton, Laurel, *Grammaticalization and lexicalization reconsidered. On the “late” use of temporal adverbs*, in: Fanego, Teresa/López-Couso, María José/Pérez de Cuellar, Javier (edd.), *English historical syntax and morphology. Selected papers from 11 ICEHL, Santiago de Compostela, 7–11 September 2000*, Amsterdam/Philadelphia, John Benjamins, 2002, 67–97.
- Brinton, Laurel/Traugott, Elizabeth Closs, *Lexicalization and language change*, Cambridge, Cambridge University Press, 2005.
- Broschart, Jürgen, *The characteristics of word classes from a crosslinguistic perspective*, in: Cruse, Alan, et al. (edd.), *Lexikology. An international handbook on the nature and structure of words and vocabularies*, vol. 1, Berlin/New York, De Gruyter, 2002, 662–674.
- Bunt, Harry, *Ensembles and the formal semantic properties of mass terms*, in: Pelletier, Francis Jeffrey (ed.), *Mass terms. Some philosophical problems*, Dordrecht, Reidel, 1979, 249–277.
- Bunt, Harry, *Mass terms and model-theoretic semantics*, Cambridge, Cambridge University Press, 1985.
- Bybee, Joan, *Language, usage and cognition*, Cambridge, Cambridge University Press, 2010.
- Callanan, Maureen/Markman, Ellen, *Principles of organization in young children’s natural language hierarchies*, *Child development* 53:4 (1982), 1093–1101.
- Cappeau, Paul/Schnedecker, Catherine, *Des “gens” bien différents à l’écrit et à l’oral. Hypothèse sur la pronominalisation des SN en “gens”*, *Verbum* 36:1 (2014), 55–74 (= 2014a).
- Cappeau, Paul/Schnedecker, Catherine, *“Gens”, “personne(s)”, “individu(s)”. Trois saisies de l’humain*, in: Neveu, Franck, et al. (edd.), *Actes du 4^{ème} Congrès Mondial de Linguistique Française, 19–23 juillet 2014*, Berlin, 2014, 3027–3040 (= 2014b).
- Carlier, Anne, *La genèse de l’article “un”*, *Langue française* 130 (2001), 177–206.
- Carlier, Anne, *Sur les premiers stades de développement de l’article partitif*, *SCOLIA* 18 (2004), 117–147.
- Carlier, Anne, *From preposition to article. The grammaticalization of the French partitive*, *Studies in language* 31:1 (2007), 1–49.
- Carlier, Anne, *Grammaticalization in progress in Old French. Indefinite articles*, in: Arteaga, Deborah (ed.), *Research on Old French. The state of the art*, Dordrecht/Heidelberg/New York/London, Springer, 2013, 45–60.

- Carlier, Anne, *The capricious evolution of the indefinite plural article “uns” and its relationship with lexical plurality in medieval French*, *Linguisticæ investigationes* 39:2 (2016), 309–334.
- Carlier, Anne/Lamiroy, Béatrice, *The grammaticalization of the prepositional partitive in Romance*, in: Luraghi, Silvia/Huumo, Tuomas (edd.), *Partitive cases and related categories*, Berlin/Boston, De Gruyter, 2014, 477–519.
- Carlier, Anne/Lamiroy, Béatrice, *The emergence of the grammatical paradigm of nominal determiners in French and in Romance. Comparative and diachronic perspectives*, *Canadian journal of linguistics* 63:2 (2018), 141–166, doi: 10.1017/cnj.2017.43.
- Chaudenson, Robert, *Bare nouns in Réunionnais Creole*, in: Baptista, Marlyse/Guéron, Jacqueline (edd.), *Noun phrases in Creole languages. A multi-faceted approach*, Amsterdam/Philadelphia, John Benjamins, 2007, 225–242.
- Cheng, Lisa/Sybesma, Rint, “*Yi-wan tang*”, “*yi-ge tang*”. *Classifiers and massifiers*, *The Tsing Hua journal of Chinese studies* 28:3 (1998), 385–412.
- Chierchia, Gennaro, *Plurality of mass nouns and the notion of “semantic parameter”*, in: Rothstein, Susan (ed.), *Events and grammar*, Dordrecht/Boston/London, Kluwer, 1998, 53–103 (= 1998a).
- Chierchia, Gennaro, *Reference to kinds across languages*, *Natural language semantics* 6:4 (1998), 339–405 (= 1998b).
- Chierchia, Gennaro, *Mass nouns, vagueness and semantic variation*, *Synthese* 174:1 (2010), 99–149.
- Cohen, Dana, *Activewear and other vaguery. A morphological perspective on aggregate-mass*, in: Moltmann, Friederike (ed.), *Mass and count linguistics, philosophy, and cognitive science*, Amsterdam/Philadelphia, John Benjamins, 2020, 37–59.
- Collin, Carl Sven Reinhold, *Étude sur le développement de sens du suffixe “-ata” (it. “-ata”, prov., esp., port. “-ada”, fr. “-ée”, “-ade”) dans les langues romanes, spécialement au point de vue du français*, Lund, Lindstedt, 1918.
- Comrie, Bernard, *Aspect. An introduction to the study of verbal aspect and related problems*, Cambridge/New York/Melbourne, Cambridge University Press, 1976.
- Comrie, Bernard, *Language universals and linguistic typology. Syntax and morphology*, Oxford, Blackwell, 1981.
- Corbett, Greville, *The agreement hierarchy*, *Journal of linguistics* 15:2 (1979), 203–224.
- Corbett, Greville, *Number*, Cambridge/New York, Cambridge University Press, 2000.
- Correia, Margarita, *Para o estudo da denominação dos conjuntos em português. Aspectos morfo-semânticos dos nomes dos colectivos*, *Filologia e lingüística portuguesa* 3 (1999), 9–21.
- Coseriu, Eugenio, *Determinación y entorno. Dos problemas de una lingüística del hablar*, *Romanistisches Jahrbuch* 7 (1955), 29–54.
- Coulmas, Florian, *Lexikalisierung von Syntagmen*, in: Schwarze, Christoph/Wunderlich, Dieter (edd.), *Handbuch der Lexikologie*, Königstein/Ts., Athenäum, 1985, 250–268.
- Croft, William, *The structure of events and the structure of language*, in: Tomasello, Michael (ed.), *The new psychology of language. Cognitive and functional approaches to language structure*, Mahwah, NJ/London, Lawrence Erlbaum Associates, 1998, 67–92.
- Croft, William, *Radical construction grammar. Syntactic theory in typological perspective*, Oxford/New York, Oxford University Press, 2001.
- Croft, William/Cruise, Alan, *Cognitive linguistics*, Cambridge, Cambridge University Press, 2004.

- Cruse, Alan, *Lexical semantics*, Cambridge, Cambridge University Press, 1986.
- Cruse, Alan, *Meaning in language. An introduction to semantics and pragmatics*, Oxford, Oxford University Press, ³2011.
- Culicover, Peter/Jackendoff, Ray, *Simpler syntax*, Oxford/New York, Oxford University Press, 2005.
- Dal, Georgette/Namer, Fiammetta, *Playful nonce-formations in French. Creativity and productivity*, in: Arndt-Lappe, Sabine, et al. (edd.), *Expanding the lexicon. Linguistic innovation, morphological productivity, and lucidity*, Berlin/Boston, De Gruyter, 2018, 203–228.
- Damourette, Jacques/Pichon, Édouard, *Des mots à la pensée. Essai de grammaire de la langue française*, Paris, Éditions d'Artrey, 1911–1927.
- Darmesteter, Arsène, *La vie des mots. Étudiée dans leurs significations*, Paris, Librairie Delagrave, 1887.
- Dębowiak, Przemysław, *La formation diminutive dans les langues romanes*, Frankfurt am Main, Peter Lang, 2014.
- Delbecque, Nicole/Verveckken, Katrien, *Conceptually-driven analogy in the grammaticalization of Spanish binominal quantifiers*, *Linguistics* 52:3 (2014), 637–684.
- Depraetere, Ilse, *On verbal concord with collective nouns in British English*, *English language and linguistics* 7:1 (2003), 85–127.
- Dessì Schmid, Sarah, *Aspektualität. Ein onomasiologisches Modell am Beispiel der romanischen Sprachen*, Berlin/Boston, De Gruyter, 2014.
- Dessì Schmid, Sarah, “Qu’est-ce que ‘le néant’?” *Grammatikalische Kategorien im Wandel. Zwischen Korpusanalyse und philologischer Arbeit*, in: Kaiser, Georg/Völker, Harald (edd.), *Philologie und Grammatik*, Berlin/Boston, De Gruyter, 2017, 225–251.
- Dessì Schmid, Sarah, *Aspectuality. An onomasiological model applied to the Romance languages*, Berlin/Boston, De Gruyter, 2019.
- Dessì Schmid, Sarah, *Zur Beziehung von progressiven Verbalperiphrasen und “states”*. *Ein erster Bericht aus Studien zu romanischen Sprachen*, *Romanistisches Jahrbuch* 72:1 (2021), 31–62.
- Detges, Ulrich, *Echt die Wahrheit sagen. Überlegungen zur Grammatikalisierung von Adverbmarkern*, *Philologie im Netz* 4 (1998), 1–29.
- Diessel, Holger, *Frequency effect in language acquisition, language use, and diachronic change*, *New ideas in psychology* 25:2 (2007), 108–127.
- Diessel, Holger, *Grammaticalization and language acquisition*, in: Narrog, Heiko/Heine, Bernd (edd.), *The Oxford handbook of grammaticalization*, Oxford/New York, Oxford University Press, 2011, 130–141.
- Diessel, Holger, *New perspectives, theories and methods. Diachronic change and language acquisition*, in: Bergs, Alexander/Brinton, Laurel (edd.), *English historical linguistics*, vol. 2, Berlin/Boston, De Gruyter, 2012, 1599–1613.
- Doetjes, Jenny Sandra, *Quantifiers and selection. On the distribution of quantifying expressions in French, Dutch and English*, Den Haag, Holland Academic Graphics, 1997.
- Doetjes, Jenny Sandra, *Count/mass distinctions across languages*, in: Maienborn, Claudia/Heusinger, Klaus von/Portner, Paul (edd.), *Semantics*, vol. 3, Berlin/New York, De Gruyter, 2011, 2559–2580.
- Dryer, Matthew, *Coding of nominal plurality*, in: Haspelmath, Martin, et al. (edd.), *The World Atlas of Language Structures*, Oxford, Oxford University Press, 2005, 138–141.

- Dubois, Jean/Dubois Charlier, Françoise, *Collectifs d'êtres vivants*, *Linx* 34:1 (1996), 125–132, doi: 10.3406/linx.1996.1422.
- Englebert, Annick, *L'article partitif. L'évolution des conditions d'emploi*, *Langue française* 109 (1996), 9–28.
- Erben, Johannes, *Zur Geschichte der deutschen Kollektiva*, in: Gipper, Helmut (ed.), *Sprache – Schlüssel zur Welt. Festschrift für Leo Weisberger*, Düsseldorf, Schwann, 1959, 221–228.
- Erben, Johannes, *Hauptaspekte der Entwicklung der Wortbildung in der Geschichte der deutschen Sprache*, in: Besch, Werner, et al. (edd.), *Sprachgeschichte*, Berlin/New York, De Gruyter, ²2003, 2525–2538.
- Fasciolo, Marco, *Des faux noms collectifs?*, in: Neveu, Franck, et al. (edd.), *Actes du 5^{ème} Congrès Mondial de Linguistique Française, 4–8 juillet 2016, Tours*, 2016, doi: 10.1051/shsconf/20162712008.
- Fernández-Ordóñez, Inés, *Dialect grammar of Spanish from the perspective of the Audible Corpus of Spoken Rural Spanish*, *Dialectologia* 3 (2009), 23–61.
- Filip, Hana, *Aspectual class and Aktionsart*, in: Maienborn, Claudia/Heusinger, Klaus von/Portner, Paul (edd.), *Semantics*, vol. 2, Berlin/New York, De Gruyter, 2011, 1186–1217.
- Fillmore, Charles/Kay, Paul/O'Connor, Mary Catherine, *Regularity and idiomacity in grammatical constructions. The case of "let alone"*, *Language* 64:3 (1988), 501–538.
- Flaux, Nelly, *Les noms collectifs*, *Analele Universității din Timișoara* 34/35 (1996–1997), 31–45.
- Flaux, Nelly, *Les noms collectifs et la prédication*, in: Forsgren, Mats/Jonasson, Kerstin/Kronning, Hans (edd.), *Prédication, assertion, information. Actes du colloque d'Uppsala en linguistique française, 6–9 juin 1996*, Uppsala, Uppsala University Library, 1998, 173–183.
- Flaux, Nelly, *À propos des noms collectifs*, *Revue de linguistique romane* 63: 251–252 (1999), 471–502, doi: 10.5169/seals-400007.
- Fleischer, Wolfgang/Barz, Irmhild, *Wortbildung in der deutschen Gegenwartssprache*, Berlin/Boston, De Gruyter, ⁴2012.
- Gaeta, Livio/Ricca, Davide, *Frequency and productivity in Italian derivation. A comparison between corpus-based and lexicographical data*, *Rivista di linguistica* 15:1 (2003), 63–98.
- Gaide, Françoise, *Les dérivés synchroniques dénominaux collectifs en latin*, *Revue de philologie, de littérature et d'histoire anciennes* 63:2 (1989), 221–228.
- Gardelle, Laure, *Are superordinates such as "furniture" and "belongings" collective nouns?*, *Linguisticae investigationes* 40:2 (2017), 150–172.
- Geeraerts, Dirk, *Reclassifying semantic change*, *Quaderni di semantica* 4 (1983), 217–240.
- Gentner, Dedre/Boroditsky, *Individuation, relativity, and early word learning*, in: Bowerman, Melissa/Levinson, Stephen (edd.), *Language acquisition and conceptual development*, Cambridge/New York, Cambridge University Press, 2001, 215–256.
- Giacalone Ramat, Anna, *Testing the boundaries of grammaticalization*, in: Giacalone Ramat, Anna/Hopper, Paul (edd.), *The limits of grammaticalization*, Amsterdam/Philadelphia, John Benjamins, 1998, 107–127.
- Gil, David, *Maltese "collective nouns". A typological perspective*, *Rivista di linguistica* 8:1 (1996), 53–87.
- Gili Gaya, Samuel, *Curso superior de sintaxis española*, Barcelona, Bibliograf, ¹¹1973.
- Gillon, Brendan, *The lexical semantics of English count and mass nouns*, in: Viegas, Evelyne (ed.), *Breadth and depth of semantic lexicons*, Dordrecht, Kluwer, 1999, 19–37.
- Givón, Talmy, *On understanding grammar. Revised edition*, Amsterdam/Philadelphia, John Benjamins, 2018.

- Goldberg, Adele, *Constructions. A Construction Grammar approach to argument structure*, Chicago/London, University of Chicago Press, 1995.
- Goldberg, Adele, *Surface generalizations. An alternative to alternations*, *Cognitive linguistics* 13:4 (2002), 327–356.
- Goldberg, Adele, *Constructions at work. The nature of generalization in language*, Oxford/New York, Oxford University Press, 2006.
- Goldberg, Adele, *Constructionist approaches*, in: Hoffmann, Thomas/Trousdale, Graeme (edd.), *The Oxford handbook of construction grammar*, Oxford/New York, Oxford University Press, 2013, 15–31.
- Gonçalves Rospantini, Maria Lígia, *A omissão do artigo em português. Um estudo contrastivo com base em jornais portugueses e brasileiros dos séculos XX e XXI*, PhD Thesis, Eberhard Karls Universität Tübingen, 2018.
- Greenberg, Joseph, *Some universals of grammar with particular reference to the order of meaningful elements*, in: Greenberg, Joseph (ed.), *Universals of language*, Cambridge, Massachusetts, MIT Press, 1963, 73–113.
- Greenberg, Joseph, *Numeral classifiers and substantival number. Problems in the genesis of a linguistic type*, in: Makkai, Adam/Becker Makkai, Valerie/Heilmann, Luigi (edd.), *Linguistics at the crossroads*, Padova/Lake Bluff, Liviana/Jupiter Press, 1977, 276–300.
- Grevisse, Maurice/Goosse, André, *Le bon usage. Grevisse langue française – Version en ligne*, Louvain-la-Neuve, de Boeck Supérieur, 2016, <https://www.grevisse.fr/etudiants-professionnels/le-bon-usage-electronique> [last access: Mar. 25th, 2022].
- Grimm, Scott, *Number and individuation*, PhD Thesis, Stanford University, 2012.
- Grimm, Scott, *Grammatical number and the scale of individuation*, *Language* 94:3 (2018), 527–574.
- Grimm, Scott/Levin, Beth, *Between count and mass. “Furniture” and other functional collectives*, talk in the framework of *Linguistics Society of America Annual Meeting (Pittsburgh, USA, January 2011)*, 2011, http://www.sas.rochester.edu/lin/sgrimm/talks/lsa2011_talk.pdf [last access: Mar. 25th, 2022].
- Grimm, Scott/Levin, Beth, *Who has more furniture? An exploration of the bases for comparison*, talk in the framework of *Mass/count in linguistics, philosophy and cognitive science conference (Paris, December 2012)*, 2012, <http://www.sas.rochester.edu/lin/sgrimm/talks/paris12mcsldes.pdf> [last access: Mar. 25th, 2022].
- Grimm, Scott/Levin, Beth, *Artifacts. Reference, countability and categorization*, talk in the framework of *Countability workshop (Düsseldorf, June 2016)*, 2016, <http://web.stanford.edu/~bclevin/artifacts-reference-countability.pdf> [last access: Mar. 25th, 2022].
- Grimm, Scott/Levin, Beth, *Artifact nouns. Reference and countability*, in: Lamont, Andrew/Tetzloff, Katerina (edd.), *Proceedings of the 47th Annual Meeting of North East Linguistic Society (NELS 47)*, Amherst, Graduate Linguistics Students Association (GLSA), 2017.
- Grinevald, Colette, *Classifiers*, in: Booij, Geert, et al. (edd.), *Morphology. An international handbook on inflection and word-formation*, vol. 2, Berlin/New York, De Gruyter, 2004, 1016–1031.
- Gross, Gaston, *Classification sémantique des collectifs humains*, *Cahiers de lexicologie* 98:1 (2011), 65–81.
- Grossmann, Maria, *Nomi collettivi*, in: Grossmann, Maria/Rainer, Franz (edd.), *La formazione delle parole in italiano*, Tübingen, Niemeyer, 2004, 244–252.
- Haspelmath, Martin, *The geometry of grammatical meaning. Semantic maps and cross-linguistic comparison*, in: Tomasello, Michael (ed.), *The new psychology of*

- language. *Cognitive and functional approaches*, vol. 2, New Jersey, Lawrence Erlbaum Associates, 2003, 211–243.
- Haspelmath, Martin, *Occurrence of nominal plurality*, in: Haspelmath, Martin, et al. (edd.), *The World Atlas of Language Structures*, Oxford, Oxford University Press, 2005, 142–145.
- Haspelmath, Martin, *Parametric versus functional explanations of syntactic universals*, in: Biberauer, Theresa (ed.), *The limits of syntactic variation*, Amsterdam/Philadelphia, John Benjamins, 2008, 75–107.
- Haspelmath, Martin/Karjus, Andres, *Explaining asymmetries in number marking. Singulatives, pluratives and usage frequency*, *Linguistics* 55:6 (2017), 1213–1235.
- Hasselrot, Bengt, *Études sur la formation diminutive dans les langues romanes*, Uppsala, Almqvist & Wiksell, 1957.
- Hasselrot, Bengt, *Étude sur la vitalité de la formation diminutive française au XX^e siècle*, Uppsala, Almqvist & Wiksell, 1972.
- Heine, Bernd/Kuteva, Tania, *World lexicon of grammaticalization*, Cambridge, Cambridge University Press, 2002.
- Herslund, Michael, *Le pluriel de l'article indéfini en ancien français*, in: Blumenthal, Peter/Tyvaert, Jean-Emmanuel (edd.), *La cognition dans le temps. Études cognitives dans le champ historique des langues et des textes*, Tübingen, Niemeyer, 2003, 75–84.
- Hilpert, Martin, *Construction grammar and its application to English*, Edinburgh, Edinburgh University Press, 2014.
- Himmelmann, Nikolaus, *Lexicalization and grammaticization. Opposite or orthogonal?*, in: Bisang, Walter/Himmelmann, Nikolaus/Wiemer, Björn (edd.), *What makes grammaticalization? A look from its fringes and its components*, Berlin/New York, De Gruyter, 2004, 21–42.
- Hohenhaus, Peter, *Lexicalization and institutionalization*, in: Štekauer, Pavol/Lieber, Rochelle (edd.), *Handbook of word-formation*, Dordrecht, Springer, 2005, 353–373.
- Hualde, José Ignacio, *Metaphony and count/mass morphology in Asturian and Cantabrian dialects*, in: Laeuffer, Christiane/Morgan, Terrell (edd.), *Theoretical analyses in Romance linguistics*, Amsterdam/Philadelphia, John Benjamins, 1992, 99–114.
- Huguin, Mathilde, *Les NH comme bases de construction morphologique. Lexèmes construits sur les noms propres de personnalités politiques françaises*, in: Schnedecker, Catherine/Mihatsch, Wiltrud (edd.), *Les noms d'humains. Théorie, méthodologie, classification*, Berlin/Boston, De Gruyter, 2018, 163–184.
- Hüning, Matthias, *Foreign word-formation in construction morphology. Verbs in “-ieren” in German*, in: Booij, Geert (ed.), *The construction of words*, Cham, Springer International Publishing, 2018, 341–371.
- Hurford, James, *Language and number. The emergence of a cognitive system*, Oxford/New York, Blackwell, 1987.
- Iliescu, Maria, *Pour une romanistique panromane. Coup d'œil sur l'étymologie de l'italien “niente” “rien”*, in: Overbeck, Anja/Schweickard, Wolfgang/Völker, Harald (edd.), *Lexikon, Varietät, Philologie. Romanistische Studien. Günter Holtus zum 65. Geburtstag*, Berlin/Boston, De Gruyter, 2011, 97–101.
- Ising, Markus, *Kognitive Ambiguität. Kollektive und duale Umkonzeptualisierungen in Grammatik und Lexik romanischer Sprachen*, Tübingen, Stauffenburg, 2019.
- Iturriz-Leza, José Luis/Skopeteas, Stavros, *Numerus*, in: Booij, Geert, et al. (edd.), *Morphology. An international handbook on inflection and word-formation*, vol. 2, Berlin/New York, De Gruyter, 2004, 1053–1066.

- Jackendoff, Ray, *Semantics and cognition*, Cambridge, Massachusetts/London, MIT Press, 1983.
- Jackendoff, Ray, *Parts and boundaries*, *Cognition* 41:1–3 (1991), 9–45 (= 1991a).
- Jackendoff, Ray, *Semantic structures*, Cambridge, Massachusetts, MIT Press, 1991 (= 1991b).
- Jackendoff, Ray, *Foundations of language. Brain, meaning, grammar, evolution*, Oxford/New York, Oxford University Press, 2002.
- Jackendoff, Ray, *Construction after construction and its theoretical challenges*, *Language* 84:1 (2008), 8–28.
- Jackendoff, Ray, *Constructions in the parallel architecture*, in: Hoffmann, Thomas/Trousdale, Graeme (ed.), *The Oxford handbook of construction grammar*, Oxford/New York, Oxford University Press, 2013, 70–92.
- Jackendoff, Ray/Audring, Jenny, *The texture of the lexicon. Relational morphology and the parallel architecture*, Oxford/New York, Oxford University Press, 2020.
- Jakobson, Roman, *Boas' view of grammatical meaning*, in: Goldschmidt, Walter (ed.), *The anthropology of Franz Boas. Essays on the centennial of his birth*, Millwood, New York, Kraus Reprint, 1959 (1974), 139–145.
- Jespersen, Otto, *A Modern English grammar on historical principles*, Copenhagen/London, Munksgaard/Allen & Unwin, 1949.
- Joosten, Frank, *Why “club” and “lingerie” do not belong together. A plea for redefining collective nouns*, in: Kleiber, Georges/Schnedecker, Catherine/Theissen, Anne (ed.), *La relation “partie-tout”*, Louvain/Paris/Dudley, Peeters, 2006, 73–88.
- Joosten, Frank, *Collective nouns, aggregate nouns, and superordinates. When “part of” and “kind of” meet*, *Linguisticæ investigationes* 33:1 (2010), 25–49.
- Joosten, Frank, et al., *Dutch collective nouns and conceptual profiling*, *Linguistics* 45:1 (2007), 85–132.
- Kabatek, Johannes, *El “singular aspectual” en la historia del español. Dos historias de un fenómeno*, in: Company Company, Concepción/Moreno de Alba, José (ed.), *Actas del VII Congreso internacional de historia de la lengua española. Mérida, Yucatán, 4–8 de septiembre de 2006*, Madrid, Arco/Libros, 2008, 745–761.
- Kastovsky, Dieter, *Wortbildung und Semantik*, Düsseldorf/Bern/München, Schwann/Bagel/Francke, 1982.
- Kiss, Tibor/Pelletier, Francis Jeffrey/Stadtfeld, Tobias, *Building a reference lexicon for countability in English*, in: Calzolari, Nicoletta (ed.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland, May 26–31, 2014*, European Language Resources Association (ELRA), 2014, http://www.lrec-conf.org/proceedings/lrec2014/pdf/361_Paper.pdf [last access: Mar. 25th, 2022].
- Kleiber, Georges, *L'article “le” générique. La généricité sur le mode massif*, Genève, Droz, 1990.
- Kleiber, Georges, *Anaphores et pronoms*, Louvain-la-Neuve, Duculot, 1994.
- Kleiber, Georges, *“Tout” et ses domaines. Sur la structure “tout” + déterminant + N*, in: Englebert, Annick (ed.), *La ligne claire. De la linguistique à la grammaire; mélanges offerts à Marc Wilmet à l'occasion de son 60^e anniversaire*, Louvain-la-Neuve, Duculot, 1998, 87–98.
- Kleiber, Georges, *Quantification universelle. “Tout les”, “chaque” et “tout”*, in: Saussure, Louis de/Rihs, Alain (ed.), *Études de sémantique et pragmatique françaises*, Bern, Peter Lang, 2012, 217–259.
- Kleineberg, Désirée, *Convergence et divergence dans l'expression de l'aspectualité nominale en français et créole réunionnais*, talk in the framework of the *Summer meeting of the Society of Pidgin and Creole Linguistics (Paris, June 28–30 2021)*, 2021.

- Kleineberg, Désirée, *Constructions as a derivational category in Spanish. Empirical evidence for a constructional schema*, in: Hennecke, Inga/Wiesinger, Evelyn (edd.), *Constructions in Spanish*, John Benjamins, accepted for publication.
- Koch, Peter, *Frame and contiguity. On the cognitive bases of metonymy and certain types of word formation*, in: Panther, Klaus-Uwe/Radden, Günter (edd.), *Metonymy in language and thought*, Amsterdam/Philadelphia, John Benjamins, 1999, 139–167.
- Koch, Peter, *Qu'est-ce que le cognitif?*, in: Blumenthal, Peter/Tyvaert, Jean-Emmanuel (edd.), *La cognition dans le temps. Études cognitives dans le champ historique des langues et des textes*, Tübingen, Niemeyer, 2003, 85–100.
- Koch, Peter, *Taxinomie et relations associatives*, in: Murguía, Adolfo (ed.), *Sens et références. Mélanges Georges Kleiber = Sinn und Referenz. Festschrift für Georges Kleiber*, Tübingen, Narr, 2005, 159–191.
- Krifka, Manfred, *Nominalreferenz, Zeitkonstitution, Aspekt, Aktionsart. Eine semantische Erklärung ihrer Interaktion*, in: Abraham, Werner/Janssen, Theo (edd.), *Tempus – Aspekt – Modus. Die lexikalischen und grammatischen Formen in den germanischen Sprachen*, Tübingen, Niemeyer, 1989, 227–258.
- Kučerová, Ivona/Moro, Anna, *On mass nouns in Romance. Semantic markedness and structural specification*, in: Armstrong, Lisa (ed.), *Actes du Congrès de l'ACL 2011*, 2011, http://homes.chass.utoronto.ca/~cla-acl/actes2011/Kucerova_and_Moro_2011.pdf [last access: Mar. 25th, 2022].
- Kuhn, Wilfried, *Formale Verfahren der Technik "Kollektion"*, in: Seiler, Hansjakob/Stachowiak, Franz Josef (edd.), *Apprehension. Das sprachliche Erfassen von Gegenständen*, 2, vol. 2, Tübingen, Narr, 1982, 55–83.
- Kuno, Susumu/Kaburaki, Etsuko, *Empathy and syntax*, *Linguistic inquiry* 8:4 (1977), 627–672.
- Kuryłowicz, Jerzy, *The evolution of grammatical categories*, *Diogenes* 13:51 (1965), 55–71.
- Lakoff, George, *Linguistic gestalts*, in: Beach, Woodford/Fox, Samuel/Philosoph, Shulamith (edd.), *Papers from the 13th regional meeting of the Chicago Linguistic Society*, Chicago, Chicago Linguistic Society, 1977, 236–287.
- Lammert, Marie, *Sémantique et cognition. Les noms collectifs*, Genève, Droz, 2010.
- Lammert, Marie, *Lexical plurals through meronymy and hyperonymy*, *Linguisticae investigationes* 39:2 (2016), 335–354.
- Lammert, Marie/Lecolle, Michelle, *Les noms collectifs en français. Une vue d'ensemble*, *Cahiers de lexicologie* 105:2 (2014), 203–222.
- Landman, Fred, *Count Nouns – Mass Nouns, Neat Nouns – Mess Nouns*, *Baltic international yearbook of cognition, logic and communication* 6 (2011), doi: 10.4148/biyclc.v6i0.1579.
- Lang, Ewald, *"Menschen" vs. "Leute". Bericht über eine semantische Expedition in den lexikalischen Nahbereich*, in: Kramer, Undine (ed.), *Lexikologisch-lexikographische Aspekte der deutschen Gegenwartssprache*, Tübingen, Niemeyer, 2000, 1–39.
- Langacker, Roland, *Foundations of cognitive grammar. Theoretical prerequisites*, Stanford, Stanford University Press, 1987 (= 1987a).
- Langacker, Roland, *Nouns and verbs*, *Language* 63:1 (1987), 53–94 (= 1987b).
- Langacker, Roland, *Concept, image, and symbol. The cognitive basis of grammar*, Berlin/New York, De Gruyter, 1991.
- Langacker, Roland, *On the continuous debate about discreteness*, *Cognitive linguistics* 17:1 (2006), 107–151.
- Langacker, Roland, *Cognitive grammar. A basic introduction*, Oxford, Oxford University Press, 2008.

- Langacker, Roland, *Entrenchment in cognitive grammar*, in: Schmid, Hans-Jörg (ed.), *Entrenchment and the psychology of language learning. How we reorganize and adapt linguistic knowledge*, Washington, DC/Boston, American Psychological Association, 2017, 39–56.
- Lasersohn, Peter, *Mass nouns and plurals*, in: Maienborn, Claudia/Heusinger, Klaus von/Portner, Paul (edd.), *Semantics*, vol. 3, Berlin/New York, De Gruyter, 2011, 1131–1153.
- Lauwers, Peter, *Les pluriels “lexicaux”. Typologie quantifiée des déficits de dénombrabilité*, *Langue française* 183 (2014), 117–132.
- Lauwers, Peter, *Les pluriels lexicaux dits “massifs” face au conditionneur universel*, *Linguisticæ investigationes* 39:2 (2016), 272–288.
- Lauwers, Peter/Lammert, Marie (edd.), *Lexical plurals and beyond. Special issue of Linguisticæ Investigationes 39 (2)*, Amsterdam/Philadelphia, John Benjamins, 2016.
- Lauwers, Peter/Vermote, Thimoteus, *La flexibilité de l’opposition massif/comptable en français et en néerlandais. Une étude contrastive*, *Syntaxe et sémantique* 15:1 (2014), 139–170, doi: 10.3917/ss.015.0139.
- Lecolle, Michelle, *Noms collectifs et méronymie*, *Cahiers de grammaire* 23 (1998), 41–65.
- Lehmann, Christian, *Grammatikalisierung und Lexikalisierung*, *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 42 (1989), 11–19.
- Lehmann, Christian, *Thoughts on grammaticalization*, Berlin, Language Science Press, 42015.
- Lehmann, Christian/Moravcsik, Edith, *Noun*, in: Booij, Geert/Lehmann, Christian/Mugdan, Joachim (edd.), *Morphology. An international handbook on inflection and word-formation*, vol. 1, Berlin/New York, De Gruyter, 2000, 732–757.
- Leisi, Ernst, *Der Wortinhalt. Seine Struktur im Deutschen und Englischen*, Heidelberg, Quelle & Meyer, 51975.
- Levin, Magnus, *Agreement with collective nouns in English*, Lund, Lund University, 2001.
- Li, Peggy/Barner, David/Huang, Becky, *Classifiers as count syntax. Individuation and measurement in the acquisition of Mandarin Chinese*, *Language learning and development* 4:4 (2008), 249–290.
- Link, Godehard, *The logical analysis of plurals and mass terms. A lattice-theoretical approach*, in: Bäuerle, Rainer/Schwarze, Christoph/Stechow, Armin von (edd.), *Meaning, use, and interpretation of language*, Berlin/New York, De Gruyter, 1983, 302–323.
- Lipka, Leonhard, *An outline of English lexicology. Lexical structure, word semantics, and word-formation*, Tübingen, Niemeyer, 1990.
- Lopes, Célia Regina, *A inserção de “a gente” no quadro pronominal do português*, Frankfurt am Main/Madrid, Vervuert/Iberoamericana, 2002.
- Loporcaro, Michele, *Gender*, in: Ledgeway, Adam/Maiden, Martin (edd.), *The Oxford guide to the Romance languages*, Oxford, Oxford University Press, 2016, 924–935.
- Loporcaro, Michele, *Gender from Latin to Romance. History, geography, typology*, Oxford, Oxford University Press, 2018.
- Lüdtke, Jens, *Romanische Wortbildung. Inhaltlich – diachronisch – synchronisch*, Tübingen, Stauffenburg, 2007.
- Luján, Eugenio, *Semantic maps and word formation. Agents, instruments, and related semantic roles*, *Linguistic discovery* 8:1 (2010), 162–175, doi: 10.1349/PS1.1537-0852.A.349.
- Lyons, John, *Structural semantics. An analysis of part of the vocabulary of Plato*, Oxford, Blackwell, 1972.
- Macnamara, John Theodore, *Names for things*, Cambridge, Massachusetts, MIT Press, 1982.

- Magni, Elisabetta, *Collective suffixes and ad hoc categories. From Latin “-ālia” to Italian “-aglia”*, *Folia linguistica historica* 39 (2018), 201–224.
- Maiden, Martin, *Number*, in: Ledgeway, Adam/Maiden, Martin (edd.), *The Oxford guide to the Romance languages*, Oxford, Oxford University Press, 2016, 697–707.
- Marchello-Nizia, Christiane, *Le français en diachronie. Douze siècles d'évolution*, Paris, Ophrys, 1999.
- Markman, Ellen, *Why superordinate category terms can be mass nouns*, *Cognition* 19:1 (1985), 31–53.
- Markman, Ellen/Horton, Marjorie/McLanahan, Alexander, *Classes and collections. Principles of organization in the learning of hierarchical relations*, *Cognition* 8:3 (1980), 227–241.
- Markman, Ellen/Seibert, Jeffrey, *Classes and collections. Internal organization and resulting holistic properties*, *Cognitive psychology* 8:4 (1976), 561–577.
- Marle, Jaap van, *The relationship between morphological productivity and frequency. A comment on Baayen's performance-oriented conception of morphological productivity*, *Yearbook of morphology* 1991 (1992), 151–163.
- Marques de Sá, Joaquim, *Pattern recognition. Concepts, methods and applications*, Berlin/Heidelberg, Springer, 2001.
- Martínez, Antonio, *La concordancia*, in: Bosque, Ignacio/Demonte, Violeta (edd.), *Gramática descriptiva de la lengua española*, vol. 2, Madrid, Espasa, 1999, 2695–2786.
- Matthews, Peter Hugoe, *Morphology*, Cambridge, Cambridge University Press, 2011.
- Mauri, Caterina, *Building and interpreting ad hoc categories. A linguistic analysis*, in: Blochowiak, Joanna, et al. (edd.), *Formal models in the study of language. Applications in interdisciplinary contexts*, Cham, Springer, 2017, 297–326.
- McCawley, James, *Lexicography and the count-mass distinction*, in: Cogen, Cathy, et al. (edd.), *Proceedings of the first annual meeting of the Berkeley Linguistics Society, February 15–17, 1975*, Berkeley, Berkeley Linguistics Society, 1975, 314–321.
- Meineke, Birgit, *Althochdeutsche “-scaff(t)“-Bildungen*, Göttingen, Vandenhoeck & Ruprecht, 1991.
- Meineke, Eckhard, *Abstraktbildungen im Althochdeutschen. Wege zu ihrer Erschließung*, Göttingen, Vandenhoeck & Ruprecht, 1994.
- Meisterfeld, Reinhard, *Numerus und Nominalaspekt. Eine Studie zur romanischen Apprehension*, Tübingen, Niemeyer, 1998.
- Merlini Barbaraesi, Lavinia, *Alterazione*, in: Grossmann, Maria/Rainer, Franz (edd.), *La formazione delle parole in italiano*, Tübingen, Niemeyer, 2004, 264–292.
- Michaelis, Laura/Lambrecht, Knud, *Toward a construction-based theory of language function. The case of nominal extraposition*, *Language* 72:2 (1996), 215–247.
- Michaux, Christine, *The collectives in French. A linguistic investigation*, *Linguisticæ investigationes* 26:1 (1992), 99–124.
- Middleton, Erica, et al., *Separating the chaff from the oats. Evidence for a conceptual distinction between count noun and mass noun aggregates*, *Journal of memory and language* 50:4 (2004), 371–394.
- Mihatsch, Wiltrud, *Wieso ist ein Kollektivum ein Kollektivum? Zentrum und Peripherie einer Kategorie am Beispiel des Spanischen*, *Philologie im Netz* 13 (2000), 39–72.
- Mihatsch, Wiltrud, *Kognitive Grundlagen lexikalischer Hierarchien. Untersucht am Beispiel des Französischen und Spanischen*, Tübingen, Niemeyer, 2006.
- Mihatsch, Wiltrud, *Collectives*, in: Müller, Peter, et al. (edd.), *Word-formation. An international handbook of the languages of Europe*, vol. 2, Berlin/Boston, De Gruyter, 2015, 1183–1195 (= 2015a).

- Mihatsch, Wiltrud, *La position taxinomique et les réseaux méronymiques des noms généraux 'être humain' français et allemands*, in: Mihatsch, Wiltrud/Schnedecker, Catherine (edd.), *Les noms d'humains – une catégorie à part?*, Stuttgart, Steiner, 2015, 85–113 (= 2015b).
- Mihatsch, Wiltrud, *Collectives, object mass nouns and individual count nouns. Nouns between lexical and inflectional plural marking*, *Linguisticæ investigationes* 39:2 (2016), 289–308.
- Mihatsch, Wiltrud, *Collectives in the Romance languages*, in: *Oxford Research Encyclopaedia of Linguistics*, 2021, <https://oxfordre.com/linguistics> [last access: Mar. 25th, 2022].
- Mihatsch, Wiltrud/Kleineberg, Désirée, *The diachronic instability of Spanish object mass nouns. A peripheral category?*, talk in the framework of the 44th Annual Conference of the German Linguistic Society (Tübingen, February 23–25, 2022), 2022.
- Mihatsch, Wiltrud/Kleineberg, Désirée, *Les pluralia tantum*, in prep. (= in prep. a).
- Mihatsch, Wiltrud/Kleineberg, Désirée, *Object mass nouns in Papiamentu in comparison to Spanish and Portuguese*, in prep. (= in prep. b).
- Millán Orozco, Antonio, *Anomalías en la concordancia del nombre en el español de la Ciudad de México*, in: Lope Blanch, Juan (ed.), *Estudios sobre el español hablado en las principales ciudades de América*, México, Universidad Nacional Autónoma de México, 1977, 125–146.
- Moltmann, Friederike, *Parts and wholes in semantics*, New York/Oxford, Oxford University Press, 1997.
- Moreno Cabrera, Juan, *On the relationship between grammaticalization and lexicalization*, in: Giacalone Ramat, Anna/Hopper, Paul (edd.), *The limits of grammaticalization*, Amsterdam/Philadelphia, John Benjamins, 1998, 211–227.
- Mulder, Walter de/Lamiroy, Béatrice, *Gradualness of grammaticalization in Romance. The position of French, Spanish and Italian*, in: Davidse, Kristin, et al. (edd.), *Grammaticalization and language change. New reflections*, Amsterdam/Philadelphia, John Benjamins, 2012, 199–226.
- Müller, Ana/Oliveira, Fátima, *Bare nominals and number in Brazilian and European Portuguese*, *Journal of Portuguese linguistics* 3 (2004), 9–36.
- Murphy, Lynne, *Knowledge of words versus knowledge about words. The conceptual basis of lexical relations*, in: Peeters, Bert (ed.), *The Lexicon-encyclopedia interface*, Amsterdam/New York, Elsevier, 2000, 317–348.
- Nicolas, David, *La distinction entre noms massifs et noms comptables. Aspects linguistiques et conceptuels*, Louvain/Paris, Peeters, 2002.
- Nixon, Graham, *Corporate-concord phenomena in English*, *Studia neophilologica* 44 (1972), 120–126.
- Norde, Muriel, *Grammaticalization versus reanalysis. The case of possessive constructions in Germanic*, in: Bentley, Delia, et al. (edd.), *Historical linguistics 1995. Selected papers from the 12th International Conference on Historical Linguistics, Manchester, August 1995*, vol. 2, Amsterdam/Philadelphia, John Benjamins, 1998, 211–222.
- Ojeda, Almerindo, *The semantics of the Italian double plural*, *Journal of semantics* 12:3 (1995), 213–237.
- Oliveira de Lima, Suzi, *The grammar of individuation and counting*, PhD Thesis, University of Massachusetts Amherst, 2014.
- Ortega, Gonzalo/Morera, Marcial, *La concordancia numérica de los colectivos. Un caso de silepsis*, *Archivum* 31–32 (1981–1982), 645–656.
- Paul, Hermann, *Prinzipien der Sprachgeschichte*, Halle (Saale), Niemeyer, ²1886.

- Pelletier, Francis Jeffry, *Non-singular reference. Some preliminaries*, *Philosophia* 5:4 (1975), 451–465.
- Pharies, David, *Tipología de los orígenes de los sufijos españoles*, *Revista de filología española* 84:1 (2004), 153–167.
- Pires de Oliveira, Roberta/Rothstein, Susan, *Bare singular noun phrases are mass in Brazilian Portuguese*, *Lingua* 121: 15 (2011), 2153–2175 (= 2011a).
- Pires de Oliveira, Roberta/Rothstein, Susan, *Two sorts of bare nouns in Brazilian Portuguese*, *Revista da ABRALIN* 10:3 (2011), 231–266, doi: 10.5380/rabl.v10i3.32352 (= 2011b).
- Plag, Ingo, *Word-formation in English*, Cambridge, Cambridge University Press, 2003.
- Poletto, Cecilia/Penello, Nicoletta, *Propagazione di suffissi derivazionali nel linguaggio giovanile*, in: Marcato, Gianna (ed.), *Giovani, lingue e dialetti*, Padova, Unipress, 2006, 73–80.
- Pomino, Natascha, *Gender and number*, in: Dufter, Andreas/Stark, Elisabeth (edd.), *Manual of Romance morphosyntax and syntax*, Berlin/Boston, De Gruyter, 2017, 691–726.
- Ponce de León, Ramón, *Análisis morfológico, semántico y lexicográfico del sufijo “-erío” en el español de México*, *Thesaurus* 58 (2016), 30–52.
- Quine, Willard van Orman, *Word and Object*, Cambridge, Massachusetts, MIT Press, 1960.
- Quirk, Randolph, *A comprehensive grammar of the English language*, New York, Longman, 1985.
- Rainer, Franz, *Spanische Wortbildungslehre*, Tübingen, Niemeyer, 1993.
- Rainer, Franz, *Polysemy in derivation*, in: Lieber, Rochelle/Štekauer, Pavol (edd.), *The Oxford handbook of derivational morphology*, Oxford/New York, Oxford University Press, 2016, 338–353.
- Rainer, Franz, *La descendencia colectiva, aumentativa e intensiva del sufijo latino “-aria” en las lenguas románicas*, *Verba* 45 (2018), 225–264, doi: 10.15304/verba.45.4129 (= 2018a).
- Rainer, Franz, *Patterns and niches in diachronic word formation. The fate of the suffix “-men” from Latin to Romance*, *Morphology* 28:4 (2018), 397–465 (= 2018b).
- Ramat, Paolo, *Thoughts on degrammaticalization*, *Linguistics* 30:3 (1992), 549–560.
- Real Academia Española/Asociación de Academias de la Lengua Española, *Nueva gramática de la lengua española. Morfología, sintaxis I*, Barcelona, Espasa, 2009.
- Real Academia Española/Asociación de Academias de la Lengua Española, *Nueva gramática de la lengua española. Fonética y fonología*, Barcelona, Espasa, 2011.
- Rijkhoff, Jan, *Nominal aspect*, *Journal of semantics* 8:4 (1991), 291–309.
- Rijkhoff, Jan, *The noun phrase*, Oxford, Oxford University Press, 2002.
- Rio-Torto, Graça Maria/Soares Rodrigues, Alexandra, *Formação de nomes*, in: Rio-Torto, Graça Maria (ed.), *Gramática derivacional do português*, Coimbra, Coimbra University Press, 2016, 135–240.
- Roedinger, Max, *Die Bedeutung des Suffixes “-ment”*, PhD Thesis, Friedrich-Wilhelm Universität zu Berlin, 1904.
- Rosch, Eleanor, et al., *Basic objects in natural categories*, *Cognitive psychology* 8:3 (1976), 382–439.
- Rosenkranz, Heinz, *Komposita auf “-ding”, “-sache”, “-werk”, “-(ge)zeug”, “-geschirr” und “-kram” im thüringischen Sprachraum*, *Beiträge zur Geschichte der deutschen Sprache und Literatur* 90 (1968), 212–248.
- Rothstein, Susan, *Counting and the mass/count distinction*, *Journal of semantics* 27:3 (2010), 343–397 (= 2010a).
- Rothstein, Susan, *Counting, measuring and the semantics of classifiers*, *Baltic international yearbook of cognition, logic and communication* 6:1 (2010), 1–42, doi: 10.4148/biyclc.v6i0.1582 (= 2010b).

- Rothstein, Susan, *Semantics for Counting and Measuring*, Cambridge, Cambridge University Press, 2017.
- Rothstein, Susan, *Counting and measuring furniture. A crosslinguistic perspective on object mass nouns*, talk in the framework of *Workshop on nominal aspect* (Tübingen, July 26, 2016), 2016.
- Rothstein, Susan, *Object mass nouns from a crosslinguistic perspective*, 2016, http://nominal-modification.de/wp-content/uploads/2015/12/ho_rothstein_feb16.pdf [last access: Mar. 25th, 2022].
- Sánchez Avedaño, Carlos, “*Para que la gente se enteren*”. *La concordancia “ad sensum” en español oral*, *Filología y lingüística* 33:2 (2007), 205–226.
- Santiago Lacuesta, Ramón/Bustos Gisbert, Eugenio, *La derivación nominal*, in: Bosque, Ignacio/Demonte, Violeta (edd.), *Gramática descriptiva de la lengua española*, vol. 3, Madrid, Espasa, 1999, 4505–4594.
- Sasse, Hans-Jürgen, *Aspect and Aktionsart. A reconciliation*, *Belgian journal of linguistics* 6 (1991), 31–45.
- Sasse, Hans-Jürgen, *Recent activity in the theory of aspect. Accomplishments, achievements, or just non-progressive state?*, *Linguistic typology* 6:2 (2002), 199–271.
- Scalise, Sergio, *Generative Morphology*, Dordrecht, Foris Publications, 1986.
- Schmid, Hans-Jörg, *Entrenchment, salience, and basic levels*, in: Geeraerts, Dirk/Cuyckens, Hubert (edd.), *The Oxford handbook of cognitive linguistics*, Oxford/New York, Oxford University Press, 2010, 117–138.
- Schmitt, Cristina/Munn, Alan, *The syntax and semantics of bare arguments in Brazilian Portuguese*, *Linguistic variation yearbook* 2 (2002), 185–216.
- Schneedecker, Catherine, “*Tout le monde*”, “*tous*”, “*(tous) les gens*”. *Relation sémantique entre des expressions dénotant la totalité /+HUM/*, in: Le Querler, Nicole/Neveu, Franck/Roussel, Emmanuelle (edd.), *Relations, connexions, dépendances. Hommage au professeur Claude Guimier*, Rennes, Presses Universitaires de Rennes, 2012, 127–155.
- Schneedecker, Catherine, *L’enrichissement du paradigme de pronoms indéfinis humains en français? Étude du processus d’évolution des SN en gens du 18^e au 19^e siècle*, in: Carlier, Anne/Goyens, Michèle/Lamiroy, Béatrice (edd.), *Le français en diachronie. Nouveaux objets et méthodes*, Bern et al., Peter Lang, 2015, 247–268.
- Schön, Ilse, *Neutrum und Kollektivum. Das Morphem “-a” im Lateinischen und Romanischen*, Innsbruck, Institut für vergleichende Sprachwissenschaft der Universität Innsbruck, 1971.
- Schulte, Marion, *The semantics of derivational morphology. A synchronic and diachronic investigation of the suffixes “-age” and “-ery” in English*, Tübingen, Narr, 2015.
- Schütze, Carson/Sprouse, Jon, *Judgement data*, in: Podesva, Robert/Sharma, Devyani (edd.), *Research methods in linguistics*, Cambridge/New York, Cambridge University Press, 2013, 27–50.
- Schwarzschild, Roger, *Stubborn distributivity, multiparticipant nouns and the mass/count distinction*, in: Oliveira de Lima, Suzi/Mullin, Kevin/Smith, Brian (edd.), *Proceedings of the 39th Meeting of the NorthEast Linguistic Society (NELS 39)*, vol. 2, Amherst, Graduate Linguistics Students Association (GLSA), 2011, 661–678.
- Seiler, Hansjakob, *Apprehension. Das sprachliche Erfassen von Gegenständen*, 3, Tübingen, Narr, 1986.
- Silverstein, Michael, *Hierarchy of features and ergativity*, in: Dixon, Robert (ed.), *Grammatical categories in Australian languages*, New Jersey, Humanities Press Inc., 1976, 112–171.
- Smith, Carlota, *The parameter of aspect*, Dordrecht/Boston/London, Kluwer, 1991.

- Smith, Peter, *The syntax of semantic agreement in English*, *Journal of linguistics* 53 (2017), 823–863, doi: 10.1017/S0022226716000360.
- Solé Solé, Elisabet, *Els noms col·lectius catalans. Descripció i reconeixement*, PhD Thesis, Universitat Pompeu Fabra, 2002.
- Soler Arechalde, María Ángeles, *El colectivo “gente” en perspectiva histórica. Problemas de concordancia*, in: Company Company, Concepción/Moreno de Alba, José (edd.), *Actas del VII Congreso internacional de historia de la lengua española. Mérida, Yucatán, 4–8 de septiembre de 2006*, Madrid, Arco/Libros, 2008, 1137–1146.
- Spelke, Elizabeth, *Initial knowledge. Six suggestions*, *Cognition* 50:1–3 (1994), 431–445.
- Stark, Elisabeth, *Gender, number, and indefinite articles*, in: Stark, Elisabeth/Leiss, Elisabeth/Abraham, Werner (edd.), *Nominal determination. Typology, context constraints, and historical emergence*, Amsterdam/Philadelphia, John Benjamins, 2007, 49–71.
- Stark, Elisabeth, *Neutrum und Individuum. Zentrale Konzepte in der Grammatik der romanischen Sprachen*, *Romanische Forschungen* 121:3 (2009), 273–319.
- Stark, Elisabeth, *Typological correlations in nominal determination in Romance*, in: Høeg Müller, Henrik/Klinge, Alex (edd.), *Essays on nominal determination. From morphology to discourse management*, Amsterdam/Philadelphia, John Benjamins, 2008, 45–63.
- Staudacher-Valliamée, Gillette, *Grammaire du créole réunionnais*, Paris, SEDES, 2004.
- Talmy, Leonard, *Lexicalization patterns. Semantic structure in lexical forms*, in: Shopen, Timothy (ed.), *Language typology and syntactic description*, vol. 3, Cambridge, Cambridge University Press, 1985, 57–149.
- Talmy, Leonard, *Path to realization. A typology of event conflation*, in: Sutton, Laurel/Johnson, Christopher/Shields, Ruth (edd.), *Proceedings of the 17th annual meeting of the Berkeley Linguistics Society, February 15–18, 1991*, Berkeley, Berkeley Linguistics Society, 1991, 480–519.
- Talmy, Leonard, *The relation of grammar to cognition*, in: Rudzka-Ostyn, Brygida (ed.), *Topics in cognitive linguistics*, Amsterdam/Philadelphia, John Benjamins, 1988, 165–205.
- Talmy, Leonard, *Toward a cognitive semantics*, vol. 2: *Typology and process in concept structuring*, Cambridge, Massachusetts/London, MIT Press, 2000.
- Tomasello, Michael, *Constructing a language. A usage-based theory of language acquisition*, Cambridge, Massachusetts/London, Harvard University Press, 2003.
- Traugott, Elizabeth Closs, *On regularity in semantic change*, *Journal of literary semantics* 14:3 (1985), 155–173.
- Traugott, Elizabeth Closs/Dasher, Richard, *Regularity in semantic change*, Cambridge, Cambridge University Press, 2002.
- Traugott, Elizabeth Closs/Trousdale, Graeme, *Constructionalization and constructional changes*, Oxford, Oxford University Press, 2013.
- Trips, Carola, *Lexical semantics and diachronic morphology. The development of “-hood”, “-dom” and “-ship” in the history of English*, Tübingen, Niemeyer, 2009.
- Tuten, Donald/Pato, Enrique/Schwarzwald, Ora, *Spanish, Astur-Leonese, Navarro-Aragonese, Judaeo-Spanish*, in: Ledgeway, Adam/Maiden, Martin (edd.), *The Oxford guide to the Romance languages*, Oxford, Oxford University Press, 2016, 382–410.
- van der Auwera, Johan, *More thoughts on degrammaticalization*, in: Wischer, Ilse/Diewald, Gabriele (edd.), *New reflections on grammaticalization. Proceedings from the International Symposium in grammaticalization, 17–19 June 1999, Potsdam*, Amsterdam/Philadelphia, John Benjamins, 2002, 19–29.
- Vendler, Zeno, *Verbs and times*, *The philosophical review* 66:2 (1957), 143–160.

- Vermote, Thimoteus/Lauwers, Peter, *L'opposition massif/comptable*, in: Apothéloz, Denis/Béguelin, Marie-José/Benzitoun, Christophe/Corminboeuf, Gilles/Deulofeu, José/Lauwers, Peter/Willems, Dominique (edd.), *Encyclopédie grammaticale du français*, 2016, <http://encyclogram.fr/> [last access: Mar. 25th, 2022].
- Vermote, Thimoteus/Lauwers, Peter/Cuypere, Ludovic de, *Transcending the lexical vs. grammatical divide regarding the mass/count distinction. Evidence from corpus studies and acceptability surveys in French and Dutch*, *Language sciences* 62 (2017), 37–51.
- Verveckken, Katrien, *Binominal quantifiers in Spanish. Conceptually-driven analogy in diachrony and synchrony*, Berlin/Boston, De Gruyter, 2015.
- Verveckken, Katrien, *Towards a constructional account of high and low frequency binominal quantifiers in Spanish*, *Cognitive linguistics* 23:2 (2012), 421–478.
- Verveckken, Katrien/Cornillie, Bert, *Un análisis cognitivo-funcional de la concordancia verbal con construcciones (pseudo)partitivas del tipo [N1 + de + N2]*, *Revue romane* 47:2 (2012), 219–257.
- Verveckken, Katrien/Delbecque, Nicole, *On the development of binominal quantifiers in Spanisch. The notion of lexical persistence revisited*, *CogniTextes* 13 (2015), doi: 10.4000/cognitextes.838.
- Vianna, Juliana, *Semelhanças e diferenças na implementação de “a gente” em variedades do português*, PhD Thesis, Universidade Federal do Rio de Janeiro, 2011.
- Vihman, Virve-Anneli/Nelson, Diane, *Effects of animacy in grammar and cognition. Introduction to special issue*, *Open linguistics* 5:1 (2019), 260–267, doi: 10.1515/opli-2019-0015.
- Vincent, Nigel, *Determination and quantification*, in: Dufter, Andreas/Stark, Elisabeth (edd.), *Manual of Romance morphosyntax and syntax*, Berlin/Boston, De Gruyter, 2017, 727–770.
- Visser, Frederic Theodoor, *An historical syntax of the English language*, Leiden, Brill, 1963.
- Wall, Albert, *Bare nominals in Brazilian Portuguese. An integral approach*, Amsterdam/Philadelphia, John Benjamins, 2017.
- Wartburg, Walther von, *Die Ausgliederung der romanischen Sprachräume*, *Zeitschrift für romanische Philologie* 56 (1936), 1–48.
- Wegerer, Martina, *Die Numeruskongruenz von Subjekt und finitem Verb im Deutschen. Untersuchung der grammatischen Entscheidungsprozeduren bei zweifelhaften Kongruenzrelationen*, PhD Thesis, Universität Wien, 2012.
- Wiedenhof, Jeroen, *A grammar of Mandarin*, Amsterdam/Philadelphia, John Benjamins, 2015.
- Wiederspiel, Brigitte, *Termes de masse et référence hétérogène*, *Le français moderne* 50:1 (1992), 46–67.
- Wierzbicka, Anna, *“Apples” are not a “kind of fruit”. The semantics of human categorization*, *American ethnologist* 11:2 (1984), 313–328, doi: 10.1525/ae.1984.11.2.02a00060.
- Wierzbicka, Anna, *Lexicography and conceptual analysis*, Ann Arbor, Karoma Publishers, 1985.
- Wierzbicka, Anna, *The semantics of grammar*, Amsterdam/Philadelphia, John Benjamins, 1988.
- Winston, Morton/Chaffin, Roger/Herrmann, Douglas, *A taxonomy of part-whole relations*, *Cognitive science* 11 (1987), 417–444.
- Wischer, Ilse, *Grammaticalization versus lexicalization. “Methinks” there is some confusion*, in: Fischer, Olga/Rosenbach, Anette/Stein, Dieter (edd.), *Pathways of change. Grammaticalization in English*, Amsterdam/Philadelphia, John Benjamins, 2000, 355–370.
- Wisniewski, Edward, *On using count nouns, mass nouns, and pluralia tantum. What counts?*, in: Pelletier, Francis Jeffrey (ed.), *Kinds, things, and stuff. Mass terms and generics*, New York/Oxford, Oxford University Press, 2010, 166–190.

- Wisniewski, Edward/Clancy, Emily/Tillman, Richard, *On different types of categories*, in: Ahn, Woo-kyoung, et al. (edd.), *Categorization inside and outside the laboratory. Essays in honor of Douglas L. Medin*, Washington, American Psychological Association, 2005, 103–126.
- Wisniewski, Edward/Imai, Mutsumi/Casey, Lyman, *On the equivalence of superordinate concepts*, *Cognition* 60:3 (1996), 269–298.
- Wisniewski, Edward/Lamb, Christopher/Middleton, Erica, *On the conceptual basis for the count and mass distinction*, *Language and cognitive processes* 18:5–6 (2003), 583–624.
- Wisniewski, Edward/Murphy, Gregory, *Superordinate and basic category names in discourse. A textual analysis*, *Discourse processes* 12 (1989), 245–261.
- Wode, Henning, *Einführung in die Psycholinguistik. Theorien, Methoden, Ergebnisse*, Ismaning, Hueber, 1988.
- Zifonun, Gisela, *Komposition (oder Halbaffigierung) zum Ausdruck von Nominalaspekt. "Schmuckstück", "Glücksfall" und "Zuckerwerk"*, in: Gaeta, Livio/Schlücker, Barbara (edd.), *Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte*, Berlin/Boston, De Gruyter, 2012, 101–133.
- Zwanenburg, Wiecher, *Correspondance between formal and semantic relations*, in: Booij, Geert/Lehmann, Christian/Mugdan, Joachim (edd.), *Morphology. An international handbook on inflection and word-formation*, vol. 1, Berlin/New York, De Gruyter, 2000, 840–850.

Subject index

- Aktionsart. *See* Verbal aspectuality
- Animacy scale 26–27, 53, 93, 97, 127, 135, 144
- Apprehension 2, 15
- Aspectual singular 32, 129
- Bare noun 114
- Bare plural 32–33, 43–44, 47, 62
 - Bare singular 32–33, 40, 49, 105, 289
- Basic level 1, 209, 264
- Classifier construction 22–26, 30, 41, 191, 247, 261, 280, 289
- Coercion 19–20, 234
- Universal grinder 19
 - Universal packager 19, 235
 - Universal sorter 19, 234
- Collectivity (concept) 2, 5–7, 11, 75, 77, 79, 85, 156, 159, 172, 179, 187–188, 193–195, 197–199, 206, 215, 227, 290–293
- Construal 2–3, 11, 13–15, 18, 20–21, 28, 35–37, 63, 142, 146, 197–198, 289
- Constructio ad sensum 50–52, 64–65, 121, 137, 142, 144, 231, 239, 272
- Cumulative reference 21, 37, 39, 44
- Definiteness scale 27
- Grammatical aspect. *See* Verbal aspectuality
- Lexical aspect. *See* Verbal aspectuality
- Lexical plural 3–4, 120
- Mass-count distinction 20, 23, 25–26, 33–34, 41, 47, 86, 96, 100–102, 105, 108, 111–114, 117, 126, 129, 132, 134, 143, 145, 190–191, 198, 205, 233–234, 237, 264, 276, 283, 289, 291
- Metonymic change 38, 67, 69, 73–75, 77–78, 155–156, 174, 189, 199, 211–212, 215–216, 220, 265, 292
- Neuter 30, 68–70, 75–76, 195, 206, 215, 241–242, 260, 278–285, 287–288
- Neutro de materia 30, 101
- Partitive article 31, 33, 40, 101, 119, 123, 190–191, 233, 236, 238, 247, 261, 268, 270, 275, 277, 280
- Pejoration 75–76, 78, 163, 167, 171–172, 179, 182, 184–185, 187–188, 195, 242, 254
- Plurale tantum 43, 62, 126, 218, 259
- Seinsart 15
- Semantic hierarchies
- Hyponymic relations 36, 54–55, 57, 59–61, 64, 120, 137–140, 142, 191, 198, 217, 222, 235, 239, 291
 - Meronymic relations 36, 38, 54–57, 59–61, 64, 120, 137, 139, 141–142, 191, 198, 217, 222, 235, 239–240, 266, 268, 291
- Set profiling 41, 47, 51, 125–126, 129, 142, 256
- Singulare tantum 42
- Singulative 68
- Stubbornly distributive predicate 27, 47–48, 51, 64, 86, 97, 120–121, 124–126, 132, 136–137, 142, 144, 197–198, 231, 239, 268, 291
- Transnumeral 42, 105, 126, 290
- Verbal aspect. *See* Verbal aspectuality
- Verbal aspectuality 16–19, 43, 199, 289

