

Indefinites in Romance and beyond

Edited by

Olga Kellert

Sebastian Lauschnus

Malte Rosemeyer

Open Romance Linguistics 7



Open Romance Linguistics

Editors:

Lorenzo Filipponio (University of Genoa)

Richard WALTEREIT (HU Berlin)

Esme Winter-Froemel (Julius-Maximilians-Universität Würzburg)

Anne C. WOLFSGRUBER (HU Berlin; FU Berlin)

In this series:

1. Kocher, Anna. Complementizers on edge: On the boundaries between syntax and pragmatics in Ibero-Romance.
2. Bullock, Barbara E. Cinzia Russi & Almeida Jacqueline Toribio (eds.). A half century of Romance linguistics: Selected proceedings of the 50th Linguistic Symposium on Romance Languages.
3. Fliessbach, Jan. The intonation of expectations: On marked declaratives, exclamatives, and discourse particles in Castilian Spanish.
4. Posio, Pekka & Peter Herbeck (eds). Referring to discourse participants in Ibero-Romance languages.
5. Pešková, Andrea. L2 Spanish and Italian intonation: Accounting for the different patterns displayed by L1 Czech and German learners.
6. Favaro, Marco. Modal particles in Italian: Adverbs of illocutionary modification and sociolinguistic variation.
7. Kellert, Olga, Sebastian Lauschus & Malte Rosemeyer (eds.). Indefinites in Romance and beyond.

ISSN (print): 2940-1097

ISSN (electronic): 2940-1100

Indefinites in Romance and beyond

Edited by

Olga Kellert

Sebastian Lauschnus

Malte Rosemeyer

Olga Kellert, Sebastian Lauschnus & Malte Rosemeyer (eds.). 2024. *Indefinites in Romance and beyond* (Open Romance Linguistics 7). Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/449>

© 2024, the authors

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: 978-3-96110-483-3 (Digital)

978-3-98554-112-6 (Hardcover)

ISSN (print): 2940-1097

ISSN (electronic): 2940-1100

DOI: 10.5281/zenodo.13644806

Source code available from www.github.com/langsci/449

Errata: paperhive.org/documents/remote?type=langsci&id=449

Cover and concept of design: Ulrike Harbort

Proofreading: Amir Ghorbanpour, Elen Le Foll, Elliott Pearl, Frank Seemann,
Guido Mensching, Hella Olbertz, Jean Nitzke, Jorina Fenner, Katja Politt, Marika
Francia, Mary Ann Walter, Rebecca Madlener, Tom Bossuyt, Valeria Quochi

Fonts: Libertinus, Arimo, DejaVu Sans Mono

Typesetting software: $\text{\LaTeX}$

Language Science Press

xHain

Grünberger Str. 16

10243 Berlin, Germany

<http://langsci-press.org>

Storage and cataloguing done by FU Berlin

Freie Universität



Berlin

Contents

Acknowledgments	iii
1 Introduction	
Olga Kellert & Malte Rosemeyer	1
2 Romanian <i>niște</i> between non-specific and specific interpretations	
Jan Davatz & Elisabeth Stark	13
3 Argentinian Spanish <i>cualunque</i> and Italian <i>qualunque</i>	
Marika Francia & Olga Kellert	45
4 Indefinite pronouns with THING and PERSON in two Ibero-Romance/Kikongo varieties: Palenquero Creole and Cabindan Portuguese	
Miguel Gutiérrez Maté	87
5 On the diachrony of Catalan indefinite <i>qualsevol</i>	
Olga Kellert & Andrés Enrique-Arias	141
6 Indefinites and quantifiers in Old Sardinian: A corpus-based study	
Guido Mensching	177
7 The Brazilian Portuguese present perfect: From nominal to verbal pluractionality	
Malte Rosemeyer & Martin Becker	221
Index	259

Acknowledgments

We would like to express our gratitude to the participants of the section “Indefinites in Romance. The limits of an instable category” at the 36th Romanistentag (Kassel) for fruitful insights and discussions about the topic of this volume. This section was organized in association with the DFG project “Quantification in Old Italian” (directed by Cecilia Poletto and Guido Mensching), of which both Olga Kellert and Sebastian Lauschnus were project members. Special thanks go to Sophie Corazzola, who provided invaluable help in the formatting of this volume. We are also grateful to the editorial team of the Open Romance Linguistics series and Lukas Heeg for careful editing and troubleshooting.

The editors of the volume are listed in alphabetic order. Olga Kellert and Malte Rosemeyer were in charge of all content-related matters. Sebastian Lauschnus and Malte Rosemeyer coordinated the formatting of the volume.

Chapter 1

Introduction

Olga Kellert

University of Göttingen

Malte Rosemeyer

Freie Universität Berlin

In this introduction to the edited volume, we establish the theoretical framework for the synchronic and diachronic study of indefinites in Romance language varieties. Due to their flexibility in interpretation, the use of Romance indefinites is highly variable and subject to dynamic processes of language change. The present volume addresses fundamental linguistic questions about language variation and change in Romance indefinites. It focuses on quantificational expressions in language varieties that have not received much attention in the previous literature, such as Old Sardinian, Argentinian Spanish, Palenquero Creole and Cabindan Portuguese, Catalan, Romanian, and others. The studies united in this volume offer new data on these processes of variation and change.

Indefinites are commonly described as linguistic elements that are used to present a referent in their scope as discourse-new (Lyons 1999: 1–15). Consequently, indefinites typically occur in presentational (1a–c) and existential (1d) contexts. They can pertain to different syntactic classes, such as indefinite articles (1a), indefinite pronouns (1b), indefinite quantifiers (1c) and bare nouns (1d) (cf. also Koch 2012).

- (1) a. There is a book on the table.
b. There is somebody outside of the house.
c. There are some snakes in the building.
d. There are snakes in Latin America.



Indefinites have received much attention in the semantic literature, due to the fact that their interpretation may differ in terms of specificity. Consider, for instance, example (2), taken from Fodor & Sag (1982); whereas in the context of (2a) the speaker appears to have had a specific referent in mind ('John') when uttering the first sentence, in (2b) she did not (cf. also von Heusinger 2002).

- (2) A student in Syntax 1 cheated on the exam.
- a. His name is John. [specific interpretation]
 - b. We are all trying to figure out who it was. [non-specific interpretation]

In his seminal work on the functions of indefinite pronouns, Haspelmath (1997: 64) established an implicational hierarchy for English that has been taken up in many subsequent studies (see, for instance, Aloni & Port 2010 for an updated version). Among other things, the hierarchy predicts that uses of indefinite pronouns in which the referent is specific and known by the speaker (3a) are more basic than pronouns in which the referent is specific and unknown (3b), which in turn are more basic than non-specific irrealis uses (3c) and free-choice pronouns (3d) (examples from Haspelmath 1997: 3).

- (3)
- a. Somebody called while you were away: guess who!
 - b. I heard something, but I couldn't tell what kind of sound it was.
 - c. Please try somewhere else.
 - d. Anybody can solve this simple problem.

Haspelmath's implicational hierarchy combines semantic and syntactic parameters (e.g. appearance in the protasis of a conditional clause or comparatives). Recent studies have extended this description of indefinites by including further features such as plurality, scalarity and modality (see Chierchia 2006, Alonso-Ovalle & Menéndez-Benito 2015, Fălăuș 2015, 2018, Kellert 2021a). Many of these studies use data from Romance languages. For instance, Chierchia (2006) demonstrates that free choice items may display a quantificational force; in the Italian example (4), it is understood that the subject referent knocked all (not just some) doors with wooden shutters. Likewise, Alonso-Ovalle & Menéndez-Benito (2015) maintain that indefinites such as Spanish *cualquiera* carry a modal meaning component; a sentence such as (5) conveys that (a) María bought a gift and (b) she could have bought any gift (the modal meaning component).

- (4) Mi sono me-sso a buss-are come un matto a
to.me be.PRS.IND.1SG start-PTCP to knock-INF like a madman to
qualsiasi porta con i battenti in legno.
any door with the shutter in wood
'I started knocking like a madman at any door with a wooden shutter.'
- (5) María compr-ó **un regalo cualquiera**.
Maria buy-PST.PFV.3SG a gift any
'María bought a random gift.'

Examples such as (4) and (5) demonstrate that that free/random choice elements such as Italian *qualsiasi*, *qualunque* and Spanish *cualquiera* differ in their distribution from English indefinites such as *any* or *random*. A more obvious example of differences between Romance languages and English is the use of articles. For instance, partitives such as French *des*, *du* etc. can be translated into English using indefinites like *some* (6a) or simply left out (6b). Here, too, we find variation between and within the various Romance languages (see Giusti & Cardinaletti 2018). As summarized in Carlier & Lamiroy (2014), the partitive is frequently optional in Italian, meaning that examples such as (6a) can be translated into Italian with or without using partitive articles (7). Given that “*de* indicates that the referent of the NP is not wholly affected by the verbal action but only partially” (Carlier 2013: 55), the difference between (7a) and (7b) might be that whereas in (7a), only a part of the available spinach was bought, (7b) is compatible with a situation in which all of the available spinach was bought.

- (6) a. J'ai achet-é **des épinards**.
I'have.PRS.1SG buy-PTCP of.the spinach
'I bought some spinach.'
- b. J'ai achet-é **du pain**.
I'have.PRS.1SG buy-PTCP of.the bread
'I bought some bread.'
- (7) a. Ho compr-ato **dello spinacio**.
have.PRS.1SG buy-PTCP of.the spinach
'I bought some spinach.'
- b. Ho compr-ato **spinacio**.
have.PRS.1SG buy-PTCP spinach
'I bought some spinach.'

Likewise, substantial variation can be found regarding the interpretation of indefinite and definite articles. To give but one example, definite articles are generally less productive in Brazilian Portuguese than in other Romance languages and English, as bare nouns can have both generic (8) and specific-definite (9) interpretations. It stands to reason that the semantics of Brazilian Portuguese definite articles is more restricted to specific-definite interpretations than in other Romance languages. Another example is variation in the expression of the personal-impersonal distinction (see the papers united in Posio & Herbeck 2023).

- (8) **Pedreiro** é preguiçoso.
 Bricklayer be.PRS.IND.3SG lazy
 ‘Bricklayers are lazy.’ (Dobrovie-Sorin & Pires de Oliveira 2008: 108)
- (9) **Quintal** é cheio de terra.
 garden be.PRS.IND.3SG full of soil
 ‘The garden is full with soil.’ (Wall 2013: 236)

Finally, definiteness and indefiniteness also play a role in the verbal domain. For instance, some Romance languages allow the use of determiners before infinitives, as in (10) (both examples are taken from Rosemeyer 2012). The parameter of definiteness appears to be crucially related to whether or not the noun phrase is interpreted as an event (10a) or a fact (10b) (Varela 1979, de Miguel 1996, Demonte & Varela 1996, Vanderschueren 2013, Schirakowski 2021), with indefinites constrained to the eventive interpretation.

- (10) a. Lo devuelve a la realidad **un cruj-ir** de hojas
 him return.PRS.IND.3SG to the reality a crackle-INF of leaves
sueitas
 loose
 ‘A crackle of loose leaves brings him back to reality.’
- b. ha sido un poco frustrante **el no haber**
 have.PRS.IND.3SG be.PTCP a bit frustrating the not have-INF
podido dilucid-ar el mecanismo responsable de la
 can.PTCP explain-INF the mechanism responsible for the
superconductividad
 superconductivity
 ‘it has been a little frustrating that we have not been able to explain the mechanism responsible for superconductivity’

Likewise, the type of a noun can have a crucial impact on the meaning of its governing verb. For instance, the type of meaning expressed by the present perfect in Indo-European languages crucially hinges on the definiteness of the direct object. Thus, whereas (11a) is likely to be interpreted as a resultative perfect, an experiential reading seems more natural for (11b). This is due to the fact that the difference between resultative and experiential readings is governed by whether or not the resultant state still holds at event time (Iatridou et al. 2003: 155–156). Whereas the use of (11a) seems plausible in a situation in which the glasses are still lost, the pluractionality of (11b), understood as reference to multiple events (Van Geenhoven 2004), appears to make such a reading implausible here. From a discourse perspective, one might argue that the resultant state of the events in (11b) is less likely to be relevant at speech time than the resultant state of the event in (11a).

- (11) a. He perd-ido **mis** gafas.
 have lost-PTCP my glasses
 ‘I have lost my glasses.’
 b. He perd-ido **muchas** gafas.
 have lost-PTCP many glasses
 ‘I have lost many glasses.’

Our brief exemplary survey has demonstrated that the semantics of Romance indefinites has been studied intensely both in the nominal and verbal domain. Far less attention has been given to the variation in terms of the differences in the use of indefinite elements between and within Romance languages. As of yet, there is no systematic comparative account of the use of indefinite elements in Romance.

To give but one example for such variation, in Brazilian Portuguese, indefinite articles are sometimes used in contexts in which no article would be used in English (see example 12 from Wall, p.c.). Notably, the indefinite determiner *uma* in (12) would be translated into Spanish using the definite determiner *la* (13).

- (12) Ao fim da tarde eu gost-o de assist-ir **uma**
 at.the end of.the afternoon I like-PRS.IND.1SG of watch-INF a
 televisão.
 television
 ‘In the evening I like to watch television.’

- (13) En la tarde me gust-a mirar la tele
in the afternoon me like-PRS.IND.3SG watch-INF the television
‘In the evening I like to watch television.’

We also document a distinct lack of studies on the historical development of Romance indefinites. Most analyses focus on processes by which indefinite articles in Romance were created or grammaticalized from transparent lexical words. For instance, Camus Bergareche & Pérez-Saldanya (2011) demonstrate how the Old Catalan adjective *diversos* (14) gradually experienced a category shift towards the use as an indefinite article (15) (examples from Camus Bergareche & Pérez-Saldanya 2011).

- (14) Los tartres són molt diversos de les altres gents, de
the tartars be.PRS.IND.3PL very different from the other people of
manera i de custums
manner and of habits
‘The Tartars are very different from other people in manners and habits’
- (15) No és dupte que diversos altres sants no hagen
not be.PRS.IND.3SG doubt that various other saints not have.PRS.SBJ.3PL
fetes moltes altres semblants e mejors abstinències
do.PTCP many other similar and greater abstinences
‘There is no question that various other saints haven’t performed many other similar and greater abstinences’

A similar process has been posited for partitives, which cannot be described as determiners in Old French (Carlier 2013), quantifiers such as Latin *aliquis* ‘some’, *nullus* ‘no’, and *nemo* ‘no one’ (Gianollo 2018), and free choice indefinites such as *qualsiasi* (Degano & Aloni 2021). However, as mentioned by Degano & Aloni (2021: 2), extremely little research has been done in this area. Most studies on the diachrony of quantifiers focus on the grammaticalization of quantifiers from transparent lexical elements into grammaticalized quantifiers (Haspelmath 1997, Verveckken 2015, Gianollo 2018, Company-Company & Pozas-Loyo 2009). Few studies concentrate on possible processes of degrammaticalization of quantifiers.

The present volume fills this lacuna in the description of Romance languages, analyzing synchronic and diachronic data from French, Spanish, Italian, Portuguese, Catalan, Romanian, Sardinian, and Palenquero. The present studies describe variation in meaning and syntactic format of indefinites in the nominal and verbal domain and offer new data as to the historical development of these constructions. The contributions to this volume address the following questions:

- Which semantic and syntactic parameters are relevant to descriptions of the distribution of Romance indefinites?
- Are their semantic properties lexically encoded or do they result from being used in a specific syntactic and/or pragmatic context?
- Which language-specific differences in the use of indefinite elements such as French *quelques*/Italian *alcuni*/Spanish *algunos* ‘some’ can be observed and how can these differences be explained?
- Can we identify historical pathways of evolution of indefinites across Romance languages? Can we document word class changes (adjective > determiner, preposition > determiner, etc.)?
- Which Romance indefinites are the result of a (de)grammaticalization process? Is the process of grammaticalization the same in all Romance languages?

The first three papers in the present volume establish comparative perspectives on the use of Romance indefinites. In their paper “Romanian *niște* between non-specific and specific interpretations”, Jan Davatz and Elisabeth Stark analyze the meaning of the Romanian indefinite *niște* ‘some’. Intriguingly, *niște* is used in similar contexts as French and Italian partitive articles, which raises the question of the categoriality of this element. Davatz and Stark use a questionnaire study to provide a fine-grained description of the semantic and syntactic properties of *niște*. Their results demonstrate that *niște* does indeed share some properties with Italian partitive articles, such as its optionality, and lack of usage with preverbal subjects in generic contexts. However, *niște* differs decisively from partitive articles in terms of its scopal properties and specificity. In particular, *niște* can have scope over negation, and its use is frequent in anaphorical contexts, where the referent has already been introduced. The authors consequently propose to analyze *niște* as a specificity marker, with properties unlike other Romance indefinite determiners. Their analysis also suggests a similarity and, possibly, competition between *niște* and the Romanian differential object marker *pe*.

The second paper “Argentinian Spanish *cualunque* and Italian *qualunque*”, by Marika Francia and Olga Kellert, addresses the question of language contact and language change. The authors discuss the difference between the Argentinian Spanish *cualunque*, which has its origin in the Italian free choice indefinite *qualunque*. They show that the Argentinian Spanish item has changed its meaning

and syntactic category to an evaluative adjective with the meaning ‘ordinary’. This change from a functional category of an indefinite into a lexical category of an adjective is particularly interesting as it attests a case of degrammaticalization. Previous research has mainly focused on which categories and elements change from one quantificational category into another (Camus Bergareche & Pérez-Saldanya 2011, Verveckken 2015) and on how quantifiers grammaticalize (Haspelmath 1997, Company-Company & Pozas-Loyo 2009). However, the question as to how quantifiers degrammaticalize has received little attention so far. The authors argue that this case of degrammaticalization is the result of a pragmatic implicature produced in specific contexts, which has caused the semantic meaning and syntactic category shift of *cualunque*.

The paper “Indefinite pronouns with *THING* and *PERSON* in two Ibero-Romance/Kikongo varieties: Palenquero Creole and Cabindan Portuguese”, by Miguel Gutiérrez Maté, is also concerned with the question of language contact. Gutiérrez Maté studies the usage of the indefinites *kusa* ‘thing’ and *hende* ‘people’ in Palenquero, a Spanish-based creole spoken in San Basilio de Palenque (Colombia), on the basis of data from his own fieldwork. His analysis shows that while *kusa* and *hende* originally expressed quantificational readings, they are evolving into indefinite pronouns. Although this grammaticalization pathway is extremely common in creoles, Gutiérrez Maté argues that the distribution of *kusa* and *hende* cannot be explained in terms of language acquisition universals. In contrast, he proposes an explanation in terms of the original substrate languages of Palenquero, Kikongo (Bantu, Sub-Saharan), where the same processes are attested. Additional evidence for this assumption is given on the basis of a corpus-based analysis of Cabindan Portuguese (Angola). Gutiérrez Maté is unable to document the grammaticalization process attested for Palenquero indefinites in Cabindan Portuguese. This result is expected, given that no Portuguese-based creole has evolved in Angola, and lends further credibility to a substratist explanation.

Language contact is also an important issue for the three papers in this collective volume that analyze historical processes of change in the domain of Romance indefinites. In “On the diachrony of Catalan indefinite *qualsevol*”, Olga Kellert and Andrés Enrique-Arias investigate the diachronic development of the Catalan indefinite *qualsevol*. They refute the existent hypothesis according to which the Catalan indefinite *qualsevol* is a loanword from a Latin indefinite and show instead that this indefinite has been grammaticalized into an indefinite from transparent lexical elements, similar to Spanish *cualquiera*, as shown by Company-Company & Pozas-Loyo (2009) and Kellert (2021a,b). The authors analyze the grammaticalization hypothesis on different levels: morphological agree-

ment, syntactic and semantic, and show that there is strong evidence for the grammaticalization hypothesis on all three levels.

The paper “Indefinites and quantifiers in Old Sardinian: A corpus-based study”, by Guido Mensching, establishes a systematic description of the inventory of Old Sardinian indefinites and quantifiers on the basis of corpus data, filling a lacuna in research on Sardinian and Romance. Mensching is particularly interested in the question of the influence of the superstratum languages Italian, Spanish and Catalan on Old Sardinian indefinites, as well as the interaction between syntactic and semantic parameters in their usage. Mensching demonstrates how a detailed analysis of the semantics and distribution of Old Sardinian indefinites can shed light on the possible origin of these forms; he argues, *contra* previous studies, that negative indefinites such as *nullu* and *perunu* ‘no (x)’ cannot be Italian loan words, whereas the quantifier *cada* must be a loan from Spanish and Catalan, and *omnia* ‘every/each’ is a Latinism. One crucial finding that supports his idea concerning negative indefinites is that while Old Sardinian was a strict negative concord language, borrowed negative indefinites frequently show a lack of negative concord in preverbal position. His analysis also uncovers historical processes of change in the system of Old Sardinian indefinites and quantifiers. In particular, the data suggest a gradual loss of agreement for the quantifier *tot(t)u* ‘all/whole’.

In the final paper of this volume, entitled “The Brazilian Portuguese present perfect: From nominal to verbal pluractionality”, Malte Rosemeyer and Martin Becker analyze the semantic change undergone by the present perfect in Brazilian Portuguese (BP), a compound tense, in a diachronic corpus of BP theater texts. On the basis of a previous analysis by Amaral & Howe (2012), the authors hypothesize that the reanalysis of the perfect occurred in transitive contexts with a direct object that is inflected for masculine and singular, but can still be interpreted as expressing plural. Later, the interpretative property of pluractionality (originally derived from the nominal complement) came to be conventionally associated with the use of the perfect, leading to the readings that are typical for today’s use of the BP perfect. Their bottom-up approach towards calculating the likelihood for a given context of the present perfect to express pluractional readings confirms this hypothesis. In addition, their analysis reveals that register variation had an important influence on this change, leading them to hypothesize that the change towards pluractional readings in the BP present perfect was facilitated by intensive contact with European Portuguese during the second half of the 19th century.

References

- Aloni, Maria & Angelika Port. 2010. Epistemic indefinites crosslinguistically. In Yelena Fainleib, Nicholas LaCara & Yangsook Park (eds.), *Proceedings of the 41st annual meeting of the North East Linguistic Society*, 29–43. Amherst, MA: GLSA.
- Alonso-Ovalle, Luis & Paula Menéndez-Benito. 2015. Epistemic indefinites: An overview. In Luis Alonso-Ovalle & Paula Menéndez-Benito (eds.), *Epistemic indefinites*, 1–30. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199665297.003.0001.
- Amaral, Patrícia & Chad Howe. 2012. Nominal and verbal plurality in the diachrony of the Portuguese present perfect. In Patricia Cabredo Hofherr & Brenda Laca (eds.), *Verbal plurality and distributivity*, 25–54. Berlin: De Gruyter.
- Camus Bergareche, Bruno & Manuel Pérez-Saldanya. 2011. From adjectives to quantifiers: Evidence from Old and Modern Catalan. *Catalan Journal of Linguistics* 10. 221–245. DOI: 10.5565/rev/catjl.62.
- Carrier, Anne. 2013. Grammaticalization in progress in Old French: Indefinite articles. In Deborah Arteaga-Capen (ed.), *Research on Old French: The state of the art*, 45–60. Dordrecht: Springer.
- Carrier, Anne & Béatrice Lamiroy. 2014. The grammaticalization of the prepositional partitive in Romance. In Silvia Luraghi & Thomas Huomo (eds.), *Partitive cases and related categories*, 477–522. Berlin, New York: De Gruyter.
- Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the “logicality” of language. *Linguistic Inquiry* 37(4). 535–590. DOI: 10.1162/ling.2006.37.4.535.
- Company-Company, Concepción & Julia Pozas-Loyo. 2009. Los indefinidos compuestos y los pronombres genérico-impersonales omne y uno. In Concepción Company-Company (ed.), *Sintaxis histórica de la lengua española (segunda parte: La frase nominal)*, 1073–1219. Mexico City: Fondo de Cultura Económica–Universidad Nacional Autónoma de México.
- Degano, Marco & Maria Aloni. 2021. Indefinites and free choice: When the past matters. *Natural Languages & Linguistic Theory* 40. 447–484.
- de Miguel, Elena. 1996. Nominal infinitives in Spanish: An aspectual constraint. *Canadian Journal of Linguistics* 41(1). 29–53.
- Demonte, Violeta & Soledad Varela. 1996. Los infinitivos nominales eventivos del español. *Signo y Seña* 7. 123–154.

- Dobrovie-Sorin, Carmen & Roberta Pires de Oliveira. 2008. Reference to kinds in Brazilian Portuguese: Definite singulars vs. bare singulars. In Atle Grønn (ed.), *Proceedings of Sinn und Bedeutung* 12, 107–121. Oslo: ILOS. DOI: 10.18148/sub/2008.v12i0.579.
- Fălăuş, Anamaria. 2018. Positive polarity indefinites? On how (not) to identify them: An exhaustification-based perspective. *Linguistics* 56(2). 301–331. DOI: 10.1515/ling-2017-0042.
- Fălăuş, Anamaria. 2015. Romanian epistemic indefinites. In Luis Alonso-Ovalle & Paula Menéndez-Benito (eds.), *Epistemic indefinites: Exploring modality beyond the verbal domain*, 60–81. Oxford: Oxford University Press.
- Fodor, Janet D. & Ivan A. Sag. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy* 5(3). 355–398. DOI: 10.1007/BF00351459.
- Gianollo, Chiara. 2018. *Indefinites between Latin and Romance*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198812661.001.0001.
- Giusti, Giuliana & Anna Cardinaletti. 2018. Indefinite determiners: Variation and optionality in Italo-Romance. In Roberta D’Alessandro & Diego Pescarini (eds.), *Advances in Italian dialectology: Sketches of Italo-Romance grammars*, vol. 1, 135–161. Amsterdam: Brill.
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
- Iatridou, Sabine, Elena Anagnostopoulou & Roumyana Izvorski. 2003. Observations about the form and meaning of the perfect. In Artemis Alexiadou, Monika Rathert & Arnim von Stechow (eds.), *Perfect explorations*, 153–204. Berlin, New York: De Gruyter. DOI: 10.1515/9783110902358.133.
- Kellert, Olga. 2021a. Free Choice Indefinites in Old and Modern Italian. *Journal of Historical Syntax* 5(1–13). 1–27. DOI: 10.18148/hs/2021.v5i1-13.82.
- Kellert, Olga. 2021b. *Variation and change of unspecific indefinites: A case study of language contact in Río de la Plata area*. Talk given at LMU on January 1st, 2021.
- Koch, Peter. 2012. Location, existence, and possession: A constructional-typological exploration. *Linguistics* 36. 533–603.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- Posio, Pekka & Peter Herbeck (eds.). 2023. *Referring to discourse participants in Ibero Romance languages* (Open Romance Linguistics 4). Berlin: Language Science Press. DOI: 10.5281/zenodo.8123675.
- Rosemeyer, Malte. 2012. On the interplay between transitivity, factivity and information structure: Spanish nominal and verbal infinitives. In Marco García García & Valeriano Bellosta von Colbe (eds.), *Aspectualidad – transitividad – referencialidad: Las lenguas románicas en contraste*, 181–209. Frankfurt: Peter Lang.

- Schirakowski, Barbara. 2021. *Nominalisierte Infinitive im Spanischen*. Berlin, New York: De Gruyter. DOI: 10.1515/9783110723359.
- Van Geenhoven, Veerle. 2004. For-adverbials, frequentative aspect, and pluractionality. *Natural Language Semantics* 12. 135–190.
- Vanderschueren, Clara. 2013. *Infinitivo y sujeto en portugués y español: Un estudio empírico de los infinitivos adverbiales con sujeto explícito*. Berlin, New York: De Gruyter. DOI: 10.1515/9783110307733.
- Varela, Soledad. 1979. Los falsos infinitivos. *Boletín de la Real Academia Española* 59. 529–551.
- Verveckken, Katrien Dora. 2015. *Binominal quantifiers in Spanish*. Berlin, München, Boston: De Gruyter. DOI: 10.1515/9783110406733.
- von Heusinger, Klaus. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics* 19. 245–276.
- Wall, Albert. 2013. The distribution of definite and specific bare nominals in Brazilian Portuguese. In Johannes Kabatek & Albert Wall (eds.), *New perspectives on bare noun phrases in Romance and beyond*, 223–253. Amsterdam, Philadelphia: John Benjamins.

Chapter 2

Romanian *niște* between non-specific and specific interpretations

Jan Davatz

University of Zurich

Elisabeth Stark

University of Zurich

Our contribution focuses on the semantics of the intriguing and highly understudied indefinite element *niște* (approximately English ‘some’) in Romanian. Given the apparent distributional similarities between *niște* and the so-called *partitive articles* (PAs) of French and Italian (both preceding either an indefinite singular mass or an indefinite plural count noun, like in French *du vin*, ‘wine’), we discuss whether and to what extent *niște* and PAs are semantically comparable. The results of a pilot study conducted in 2018 with 33 Romanian native speakers show, first, that similarly to the Italian and unlike the French PAs, the use of *niște* is in no context obligatory, but rather strongly connected to individual preferences. Second, *niște* is never used with preverbal subjects in generic contexts, behaving thus similarly to the Italian PA (facts for the French PA are unclear). Third, and most intriguingly, next to narrow scope readings of *niște* with respect to negation, wide scope is not excluded, neither with plural count nor with mass nouns. This last behavior is unattested for both the Italian and the French singular PA (and related Gallo-Romance languages such as Franco-Provençal). Fourth, two specificity-related properties of *niște* seem to depend on the task: While the speakers clearly prefer a noun determined by *niște* over a bare noun in the case of epistemic specificity, they do not produce it actively in the translations. A similar observation can be made for the combination with the (specificity marking) DOM-marker *pe*, which is considered grammatical by the majority, but apparently not preferred in active production. In sum, its scope properties and its apparently strong connection to epistemic specificity make *niște* an element of its own, not comparable with any other Romance indefinite determiner.



1 Introduction

The Romanian indefinite element *niște* is often associated with the so-called *partitive articles* (PA in the following) of Romance languages such as French or Italian, as it seems to have a similar distribution:

- (1) a. French
Je bois *(du) vin.
- b. Italian
Bevo (del) vino.
- c. Romanian
Beau (*niște*) vin.
'I drink (some) wine.'
- (2) a. French
Je vois *(des) enfants.
- b. Italian
Vedo (dei) bambini.
- c. Romanian
Văd (*pe niște*) tineri.
'I see (some) children.'

As we can see from the examples in (1) and (2), only French does not allow bare arguments, not even for singular mass nouns (cf. 1a). Semantically, PAs in French and Italian are described as having a narrow scope/non-specific reading in the singular, but wide scope and specific readings available for the plural (cf. Ihsane 2008; Cardinaletti & Giusti 2016).

In this contribution, which is based on a collection of experimental data (fieldwork, translation, interpretation, grammaticality judgments, cf. Cornips & Poletto 2005), we aim at a detailed semantic description of *niște*, also in a comparative (Romance) perspective.

The paper is structured as follows: After a short summary of semantic descriptions of *niște* available in the literature (§2), including an introduction to several different notions of specificity and our working definition and a glance at the Romanian DOM-marker *pe*, we present our methodology and data in §3. §4 summarizes the main results of our fieldwork study, with a focus on the behavior of *niște* with respect to specificity. These results are discussed in §5, especially in §5.2 and §5.3, before a short conclusion in §6, emphasizing the idiosyncratic character of *niște* in a pan-Romance perspective.

2 State of the art

This section is subdivided in five subsections, building the basis for the remainder of the article. After a short presentation of the etymology of *niște* and its first uses in Old Romanian documents in §2.1, main insights from the spare existing contributions concentrating on its semantics are summed up in §2.2. §2.3 and §2.4 introduce the notion of *specificity* and the DOM-marker *pe* with its specificity-related properties, both crucial to the understanding of the semantics of *niște*. §2.5 discusses the research questions underlying this article.

2.1 The origin of *niște*

The etymology of *niște* is most likely to be found in the Latin expression *NESCIO QUID* ‘I don’t know what’.¹ The use of *niște* is attested, often also under the form *nește*, already in the first Old Romanian documents from the 16/17th century (cf. Stan 2006). From the first attestations onwards, *niște* is used both with plural count (3) and mass nouns (4), the second use being, however, less frequent (Pană Dindelegan 2016: 299, 354)

- (3) Old Romanian (Pană Dindelegan 2016: 299)

Au venit nește boiari.
have.3PL come.PTCP NIȘTE boyar.PL
‘Some boyars came.’

- (4) Old Romanian (Pană Dindelegan 2016: 354)

nește oloi
NIȘTE oil
‘some oil’

In addition to these two uses, which still exist in present-day Romanian, another use of *niște* is attested. According to Pană Dindelegan (2016: 354), this use is excluded in Modern Romanian. Consider example (5), where *niște* precedes a numeral with a plural count noun:

- (5) Old Romanian (Pană Dindelegan 2016: 354)

Niște trei voinici
NIȘTE three heroic.PL
‘some three heroes’

¹An alternative etymology which has been put forward is the evolution from abbreviated Romanian *nu știu ce* ‘I don’t know what’, i.e. *nuș’ ce*, to *niște* (Stan 2006: 200).

2.2 Semantic descriptions of *niște*

First of all, following the existing literature, we have to distinguish between the use of *niște* in combination with mass nouns (conjugated as singular) from the cases where *niște* precedes a plural count noun. For reasons of convenience, we are going to refer to the first one as *niște_{SG}* and to the second one as *niște_{PL}*. As for the semantics of *niște_{SG}*, we find a first approximation of its semantics in Romalo (2005):

Pentru a exprima aproximarea nonspecifică, se folosește în contextul substantivelor masive [...] *niște*, intrând în opoziție semantică cu *mult*, *puțin*, care exprimă aproximarea specifică. (Romalo 2005: 261)

[In order to express non-specific approximation, in the context of mass nouns *niște* is used, which enters a semantic opposition with *mult*, *puțin*, which express specific approximation.]

According to the author, *niște_{SG}* thus expresses “non-specific approximation”, which is semantically opposed to other quantifiers expressing “specific approximation”, like *mult* ‘much’ or *puțin* ‘little’. We interpret this in terms of the difference between specific and non-specific quantification; contrary to *mult* (‘much’ = big amount) vs. *puțin* (‘little’ = small amount), *niște* does not inform about the size of the amount of the substance at issue.² Following this description, the semantic contribution of *niște_{SG}* seems to be restricted to the mere assertion of a certain amount which, however, is left unspecified. This seems to be fine with concrete mass nouns, but more difficult to conceive of with abstract nouns like *talent* ‘talent’ (cf. Nedelcu 2003: 2).

Concerning *niște_{PL}*, Dobrovie-Sorin (2013: 65) states the following in their reference grammar:

[...] with plural count nouns, *niște* ‘some’ may introduce an individualized plural entity, distinguishable from other plural entities of the same kind.

The same authors provide two examples to illustrate the difference with respect to *niște_{SG}*:

²See, however, Davatz & Stark (2021), where we show that the speakers seem to have a clear idea of the amount denoted by *niște*. According to our findings, *niște* denotes a small quantity and can be situated on the continuum between *mult* (‘much’) and *puțin* (‘little’).

(6) Modern Romanian

Maria a văzut niște filme, iar Ion a văzut
 Mary have.3SG see.PTCP NIȘTE movie.PL but John have.PRS.3SG see.PTCP
 altele.
 other.PL

‘Mary has seen some movies, but John has seen others.’

(7) Modern Romanian

#Maria a băut niște vin, iar Ion a
 Maria have.3SG drink.PTCP NIȘTE wine but John have.PRS.3SG
 băut altul.
 drink.PTCP other

‘Mary has drunk some wine, but John has drunk another.’³

In von Heusinger’s (2002) terms, this seems to be indicative of *referential* or *epistemic specificity* for *niște_{PL}*, i.e. the referents of the respective DP have already been introduced in the discourse universe and/or are known by the speaker.

Next to its property of introducing an individualized plural entity, *niște_{PL}* also seems to be – at least to a certain degree – compatible with generic readings, even in preverbal position. Nedelcu (2009: 207) gives the following example:

(8) Modern Romanian

Nu pot să cred că acei doi sunt
 NEG can.PRS.1SG COMP believe.PRS.1SG COMP DEM.MPL two be.PRS.3PL
 milionari. Niște milionari nu călătoresc la clasa
 millionaire.PL NIȘTE millionaire.PL NEG travel.PRS.3PL at class.ART
 a doua.
 second

‘I can’t believe those two are millionaires. Millionaires don’t travel economy class.’

This use is, however, not uncontroversial in the literature. According to Avram (1986: 82), in the singular both the definite and the indefinite article can be used with a generic reading, like in many Indo-European languages, whereas in the plural, only the definite article can be used with generic interpretation, *niște_{PL}* being excluded from the subject position in generic contexts. Its use in (8) might be explained by the fact that the respective DP (*niște milionari*) does not refer to

³The sentence is grammatically fine, but the contrast exemplified in (6) is much harder to obtain with mass nouns.

millionaires in general, but rather to a certain group (of millionaires) (cf. Nedelcu 2009: 207).

Furthermore, unlike bare nouns, i.e. nouns without any determiner, nouns preceded by *niște*_{PL} can also have wide scope with respect to intensional predicates. Consider the two readings (a) and (b) of (9) illustrating narrow scope and wide scope of the DP preceded by *niște*_{PL}, respectively (example taken from Dobrovie-Sorin 2013: 63):

- (9) Modern Romanian
 Maria crede că Petre a furat niște cărți.
 Mary believe.PRS.3SG COMP Peter have.PRS.3SG steal.PTCP NIȘTE book.PL
 a. ‘Mary believes that Peter stole books (no matter which ones).’
 b. ‘There are books of which Mary believes that Peter stole them.’

2.3 The notion of specificity

As could easily be seen in example (9), DPs introduced by *niște* seem to be able to show *scopal specificity*, i.e. specificity that is induced by the interaction of an indefinite with other operators in the sentence, e.g. the predicate ‘believe’ in the case of (9), universal quantifiers, negation etc. This *scopal specificity* is, however, only one out of four different concepts of *specificity* described in the semantic literature.

Next to *scopal specificity*, which is conditioned by the presence of a variable-binding operator, there is the so-called *epistemic specificity*, which can be best described as a specificity dependent on the speaker’s knowledge. By *epistemically specific* DPs we understand (indefinite) DPs which are “inherently” referential because their referents are known by the speaker at the time of uttering the sentence. Consider (10) for an English example, given by von Heusinger (2002: 260), where the student referred to as “a student” is known to the speaker:

- (10) English
 A student in Syntax 1 cheated on the exam. His name is John.

In addition to *scopal* and *epistemic specificity*, the literature further lists *partitive specificity* and *relative specificity* as types of specificity (cf. Farkas 1995 and von Heusinger 2002). As for *partitive specificity*, it induces a strong (presuppositional) interpretation of an indefinite DP (both the existence of a superset to which the partitive specific DPs belong, and the existence of their referents, is presupposed). *Partitively specific* indefinite DPs always have wide scope with

respect to other operators and can be considered the equivalent of what Mil-sark (1974) called *strong indefinites*. The sentence in (11) shows one such example, again taken from von Heusinger (2002: 260):

(11) English

Some ghosts live in the pantry, others live in the kitchen.

As clearly illustrated by the sentence in (11), this partitive interpretation is most easily induced by using a “complementary” pronoun such as ‘others’.⁴ In the absence of such a pronoun, the most natural interpretation would be a weak (existential) indefinite, the strong one being limited to cases of non-canonical intonation (SOME *ghosts live in the pantry*). Importantly, the referent of the indefinite DP is presuppositional, i.e. its existence cannot be negated. However, it does not necessarily have to be known to the speaker.

The last type of specificity discussed in the literature is the so-called *relative specificity*. Relatively specific indefinite DPs are neither wide scope nor referential, but still “specific” as they refer independently from the matrix predicate (cf. von Heusinger 2002: 262):

(12) English

James said that George met a certain student of his.

Note that in (12) there is “referential co-variation” of the variable introduced by the specific indefinite DP *a certain student* and the proper name *George*: The value for *a certain student* is dependent on George, for whom the referent is necessarily specific. On the contrary, James does not have to be able to identify the student.

Against the wide discussion of such examples and consensus in the literature about the existence of these four different types of specificity, we would like to follow the unifying approach of von Heusinger (2002), according to whom *specificity* is best understood as *referential anchoring*:

Specificity indicates that an expression is referentially anchored to another object in the discourse. “Referentially anchored” means that the referent of the specific DP is functionally dependent on the referent of another expression.
(von Heusinger 2002: 268)

This conception enables us to reconcile three of the four types of specificity, as von Heusinger clearly illustrates with the following example (2002: 269):

⁴Note that *partitive specificity* has to be distinguished from cases like Example (6), where the two DPs do not belong to a common discourse-given set.

(13) English

- a. Bill gave each student a (certain) task_{speaker} to work on.
- b. Bill gave each student a (certain) task_{Bill} to work on.
- c. Bill gave each student (x) a (certain) task_x to work on.

As these different interpretations show, an indefinite DP, or more precisely its index, can be linked to different established indices resulting in different types of specificity. The first interpretation in (13a) corresponds to the epistemic specificity, where the indefinite DP *a (certain) task* is anchored to the speaker index and thus completely independent of other elements in the sentence. In (13b), it is anchored to the subject index, i.e. in this case there is a certain task that Bill gave to each student but which may be obscure to the speaker. Its interpretation corresponds thus to a relatively specific DP. In the third reading, the specific DP is anchored to the quantified DP *each student* and the task varies thus from student to student. Accordingly, we are dealing with *scopal specificity* in this case.

2.4 The DOM-marker *pe* as a specificity-marking element

An element strongly connected to the notion of specificity is the Romanian DOM-marker *pe*, which is presented briefly in this section. Knowing its properties helps to understand its interaction with *niște*_{PL}, which seems, as we have seen, to be linked to (epistemic) specificity as well (cf. §2.2). As the interplay (and grammaticality) of *pe* and *niște*_{PL} is undescribed in the literature, our pilot study involves several examples of a combination of these two elements, allowing us to draw a clearer picture of the semantics of *niște*.

Specificity is a grammatically relevant feature in Romanian, where the DOM-marker *pe* is highly sensitive to the specificity of the referent. Put differently, non-specificity blocks the appearance of *pe* (Chiriacescu & von Heusinger 2010: 303; Stark 2011: 42). Consider (14a) with a specific DP marked by *pe* vs. (14b) with a non-specific DP, which has to be unmarked (examples taken from Chiriacescu & von Heusinger 2009: 303–304).⁵ Note that the marker *pe* is “tightly related to Clitic Doubling” (Hill & Mardale 2017: 393), cases of *pe* without a co-occurring clitic being considered “marginal” (Chiriacescu & von Heusinger 2009: 7). In (14a), the feminine singular clitic *o* doubles the DP *o secretară* ‘a secretary’, whereas clitic doubling is excluded in the absence of *pe*, as (14b) shows:

⁵The referent of the indefinite DP in (14b) could also be interpreted specifically, but a non-specific interpretation of (14a) is excluded.

(14) Modern Romanian

- a. Ion o caută pe o secretară.
John CL search.PRS.3SG *pe* a secretary
'John is looking for a secretary (whom he knows).'
- b. Ion caută o secretară.
John search.PRS.3SG a secretary
'John is looking for a secretary.'

The contrast exemplified in (14) thus concerns scopal specificity, i.e. the fact that the indefinite DP can scope over the predicate *search*. In fact, it is not possible for a *pe*-marked direct object to scope under extensional/intensional operators. Note that in von Heusinger's framework, the index of the secretary in (14a) is referentially anchored to the index of the subject, John.

As for transparent contexts without any operators, epistemic specificity may occasionally trigger *pe*-marking, too. Consider (15), where according to von Heusinger & Chiriacescu (2013: 443) both the version with and without the marker *pe* could receive the continuation 'I do not know the friend' or 'I do know the friend':

(15) Modern Romanian

- Petru (l-) a vizitat (pe) un prieten.
Peter CL have.PRS.3SG visit.PTCP *pe* a friend
'Peter visited a friend.'

To put it in von Heusinger's framework, the *pe*-marked indefinite direct object can be anchored to the speaker of the utterance, but does not have to be. The same holds for the unmarked indefinite DP.⁶

We can thus conclude that "if an indefinite noun phrase is *pe*-marked, it must be scopally [in combination with extensional operators] or referentially [in combination with intensional operators] specific" (Chiriacescu & von Heusinger 2010: 305). It does not have to be necessarily epistemically specific.

⁶If we want to follow von Heusinger & Chiriacescu (2013: 443), the (subtle) difference between the two forms can be explained by introducing a discourse-based parameter. According to them, "*pe*-marking signals a higher referential persistence". However, similarly to the observations concerning the specificity effects of *pe*, "the lack of *pe*-marking does not necessarily signal a lower level of referential persistence" (Chiriacescu & von Heusinger 2010: 315).

2.5 Research questions

This article seeks to contribute to the description of the semantics of *niște_{PL}* by closing several gaps identified in the literature and asking new questions. First and foremost, we want to further explore the difference between a bare plural count noun and a DP introduced by *niște_{PL}* (in analogy to the discussion in Italian, where DPs introduced by a “partitive article” are semantically opposed to bare nouns). Based on the observation that *niște_{PL}* “may introduce an individualized plural entity” (cf. §2.2), we hypothesize that (epistemic) specificity might be a crucial factor when it comes to its use. In other words: (epistemically) specific DPs might tend to be marked by *niște_{PL}*. As the same holds true for DPs marked by the DOM-marker *pe* (cf. §2.4), the question has to be asked whether a combination of the two elements is a priori possible and, if yes, whether this has any effects on the interpretation of the DP.

Second, we want to address the controversial question of *niște_{PL}* in generic contexts, and, third, complete the descriptions regarding the scope properties of *niște* (cf. §2.2), which lack two fundamental aspects: (i) How does *niște_{PL}* behave with respect to other quantifiers, and (ii) how does *niște* behave with respect to negation? Finding answers to these three questions will allow us to compare *niște* with the French and Italian “partitive article” from a semantic point of view.

3 Methodology and database

The following section will present the methodology we used to collect our data (§3.1) and the data on which our findings are based (§3.2).

3.1 Methodology

In order to gather data that could be used for the description of the semantic properties of *niște*, we designed a questionnaire consisting of four different tasks: (i) translation, (ii) interpretation, (iii) preference and (iv) grammaticality judgments. The tasks had to be done by the speakers in the order just mentioned, avoiding thus a bias in the translation task. (i) was composed of 31 German sentences which had to be translated into Romanian. The 31 sentences contained, all in all, 9 mass nouns and 17 count nouns without any kind of determiner (some sentences containing both types of nominals) which in principle could be translated either by a bare noun or a DP introduced by *niște*. Two mass nouns (*etwas Kürbis* ‘some squash’, *ein wenig Wein* ‘some wine’) and 3 count nouns (*ein paar wenige Fehler* ‘some few mistakes’, *einige Leute* ‘some people’, *ein paar wenige Krümel* ‘some

few breadcrumbs’) were introduced by one or more quantifiers, which could be translated by *niște* or other quantifiers. Thirty-one nominals could thus in principle be translated by a DP introduced by *niște*. Additionally, 2 mass nouns which were part of a partitive construction (*von diesem Kuchen* ‘of this cake’, *von seinem Bier* ‘of his beer’) and one count noun introduced by a (colloquially modified) numeral (*so drei Idioten* ‘(some) three idiots’) were added as fillers.

Both mass nouns and count nouns were tested in direct object position, as prepositional and presentational complements, with stage-level predicates (expressing transitory properties, cf. Carlson 1977) and in generic and negative contexts. The goal of the translation task was thus to identify possible syntactic contexts where *niște* is obligatory.

The interpretation task (ii) consisted of 6 different Romanian sentences whose respective interpretations had to be indicated by the participants.⁷

The preference task (iii) was composed of 9 different sentences: 6 sentences contrasted the use of a bare noun, i.e. a noun without any kind of determiner, and the use of a DP introduced by *niște*. Half of these sentences contained a mass noun in different syntactic contexts (preverbal subject of a generic sentence, presentational complement and direct object), half of them contained a count noun (two times in direct object position, once as a presentational complement). The speakers had to indicate whether they prefer the version with *niște* or the one without. One sentence focused on the presence vs. absence of the DOM-marker *pe* (and clitic doubling) in combination with *niște*, i.e., *niște* was present in all three versions of the sentence. In addition to these two types, there were two sentences testing word order properties, which are not discussed in the remainder of this article. The preference task was meant to complement the translation task and check whether the participants behave according to their active productions.

Test set IV consisted of 40 sentences containing an occurrence of *niște*, whose grammaticality had to be judged by the speakers on a reduced Likert scale from 0 (= “I don’t understand the sentence”) to 3 (= “the sentence is well formed”). Reducing the scale to 4 values should prevent the speakers from spending too much time on thinking about slight and – for our purposes – irrelevant differences regarding the “usualness” of a sentence and allow them to focus on the difference between grammatical and ungrammatical. However, it seemed important to us to give them the possibility to indicate in case they had not understood the sentence (which could point to its ungrammaticality) or if a sentence is grammatical,

⁷Four out of these 6 sentences, all of them with a binary choice for the participants regarding their interpretation, served to test the scope properties of *niște* (cf. §4.4). Additionally, one sentence focused on the collective vs. distributive interpretation of *niște_{pl}* (cf. footnote 17) and one on the quantitative interpretation of *niște_{sg}* (cf. Davatz & Stark 2021).

but (very) unusual. One of our main concerns here was the possibility of a combination of *niște* with the DOM-marker *pe* (cf. §2.4). The participants could take the time they needed to answer the questionnaire in written form.

3.2 Data

The questionnaire presented above was used in fieldwork in March 2018 at the Babeș-Bolyai University in Cluj (Romania). All in all, we have data from 61 students of German philology, 32 of them Romanian native speakers and 29 Hungarian native speakers. In the remainder of this article, only the former will be considered.⁸

The 32 questionnaires from the Romanian native speakers contain altogether 908 valid translations of the 29 relevant nominals in the translation task, of which 44 are introduced by *niște*. For the task regarding the preference of the presence/absence of *niște*, which in principle should have generated a total of 192 responses (32×6), we count 199 responses, 126 of which contain *niște*. This is due to the fact that various speakers left out some examples, while others accepted both versions, especially with the ones with presentational constructions: *Nouă ne place când este (niște) zăpadă* ‘We like it when there is (some) snow’ and *Sunt (niște) oameni pe lumea asta care nu te-ar ajuta niciodată* ‘There are (some) people in this world who would never help you’. Counting only the examples where one single version is indicated as correct, there are 110 occurrences of *niște* and 57 occurrences of bare nouns.

As far as the interpretation task and the grammaticality judgment are concerned, we have valid and unambiguous data from all speakers.

4 Results

This section presents the findings concerning the semantic properties of *niște* that result from our study. In §4.1 we first show some general insights from the translation task regarding the use of *niște* in active production. §4.2 is concerned with the results regarding the use of *niște* with generic nominals in the translation task. The following §4.3. treats the use of *niște* with specific nominals and is divided into §4.3.1, focusing on the results of the preference task, and §4.3.2, showing the findings from the translation task. §4.4 presents the scope properties of *niște* resulting from the interpretation task. Finally, §4.5 concerning the

⁸For a discussion of the results of the Hungarian participants see Davatz (2018).

combination of *niște* with the DOM-marker *pe* is again divided in two subsubsections §4.5.1 and §4.5.2, which show the results of the preference task and the grammaticality judgment task, respectively.

4.1 Generalities

The first, very general, but nonetheless important and new finding resulting from the translation task is that there does not seem to be any syntactic context in which the use of *niște* is obligatory. *Niște* is used only in 44 out of the relevant 908 translations in the respective task, which amounts to 5%. Interestingly, there is no difference regarding the frequency of its use between (singular) mass nouns and (plural) count nouns: With count nouns, *niște* is used in 31 out of 620 translations (= 5%), with mass nouns in 13 out of 288 possible cases, which amounts to exactly the same ratio. The results show thus that in the vast majority of the cases the participants (i) prefer a bare noun to the use of *niște* and (ii) would rather use a quantifier different from *niște* in active production. As far as its apparent optionality is concerned, it resembles *prima facie* the PA of Standard Italian, which use is traditionally said to be non-obligatory as well.

These preliminary findings are, however, not confined to the simple observation that *niște* is never obligatory, but show furthermore that its use seems to be strongly connected to individual preferences. There are two crucial numbers supporting this observation: (i) Only 19 out of 32 speakers used *niște* at least once in their translations, which means that more than one third of the speakers did not make use of it at all, and (ii) only 5 out of the 32 speakers (= 16%) are responsible for 27 of the 44 occurrences (= 61%). Since the only controlled sociolinguistic variable was the education of the speakers, other extralinguistic variables such as diastratic or diatopic factors might play a role in the use of *niște*.⁹ See Table 1, which shows the distribution of *niște* over the different sentences among the 9 speakers using it more than just once.

⁹With plural abstract nouns, the use of *niște* is generally not recommended (Avram 1986: 79) and to be understood as a stylistic means to express irony or, in some cases, admiration (Nedelcu 2003: 4–6). However, according to Nedelcu (2003: 5), there is a tendency in colloquial registers, and even in the media, that *niște* is used also with plural abstract nouns in stylistically unmarked contexts, being reduced to its function of a mere indefinite determiner. It is thus by no means excluded that the use of *niște* could be influenced either by diastratic or stylistic factors also with plural concrete nouns.

Table 1: Usages of *nişte* over the different sentences among the speakers using it more than once (translation task) (cf. Davatz 2018: 39)

Speaker	Number of the sentence in the questionnaire												Total
	3	5	6	12	13	17	18	25	27	28	29	30	
1	✓	✓		✓		✓	✓			✓	✓	✓	8
2	✓	✓		✓					✓			✓	5
3	✓						✓		✓		✓	✓	5
4		✓			✓	✓				✓		✓	5
5	✓							✓	✓			✓	4
6		✓								✓			2
7	✓				✓								2
8	✓				✓								2
9	✓		✓										2
	7	4	1	2	3	2	2	1	3	3	2	5	35

4.2 Use of *nişte* with generic nominals

The results concerning the relevant input sentence in the translation task show clearly that the use of *nişte* does not seem to be compatible with a generically interpreted nominal. In 27 out of the 28 valid translations we find the definite article, and there is no single translation making use of *nişte*. Consider the German input sentence in (16) and in (17) an example of a typical translation by the informants:¹⁰

(16) German

Ich kann nicht glauben, dass diese zwei Millionäre sind. Millionäre reisen nicht in der zweiten Klasse.

‘I can’t believe those two are millionaires. Millionaires don’t travel economy class.’

¹⁰To avoid overloading the questionnaire of this pilot study, we did not test every context using all the different tasks. The results of the translation task are thus the only results we have concerning the possibility to use *nişte* in combination with generic plural nominals. The preference task includes, however, a generically interpreted singular nominal in a sentence translated as “Rice is more nourishing than polenta”. The results show that *nişte* seems to be possible in such cases but that the use of the definite article is strongly preferred. Furthermore, *nişte* seems to indicate rather a (small) quantity than pure indefiniteness (cf. Davatz 2018, Davatz & Stark 2021).

(17) Modern Romanian

Nu pot să cred că ăștia doi sunt
 NEG can.PRS.1SG COMP believe.PRS.1SG COMP DEM.MPL two be.PRS.3PL
 milionari. Milionarii nu călătoresc cu clasa
 millionaire.PL millionaire.PL.ART NEG travel.PRS.3PL with class.ART
 a doua.
 second

4.3 Use of *niște* with specific nominals

In order to allow for a solid comparison between the results of the different tasks, the results concerning the use of *niște* with specific nominals are subdivided in a subsection presenting the results of the preference task and one dealing with the findings from the translation task. This also allows us to highlight the importance of the type of task the informants are given.

4.3.1 Preference task

There are two different preference tasks in the questionnaire concerning the use of *niște* with specific nominals. The first one contrasts a bare noun and a noun phrase preceded by *niște* in the context of an epistemically specific object. (18) illustrates the two options between which the speakers had to decide:

(18) Modern Romanian

Mama a întâlnit (niște) membri ai
 mother.ART have.PRS.3SG meet.PTCP NIȘTE member.PL AGR
 parlamentului: Șerban și Ioan.
 parliament.ART.GEN Șerban and Ioan
 ‘Mum has met some members of the parliament: șerban and Ioan.’

The numbers show a predominant preference for *niște* with such specific object nominals, as Table 2 clearly illustrates.

Table 2: Preferences with respect to presence/absence of *niște* with a specific object nominal

Bare noun	<i>niște</i>
4 (12%)	29 (88%)

The second preference input item contrasts again a bare noun with a noun preceded by *niște*, but in this case, the context is slightly different. The object noun phrase does not necessarily denote a specific referent, but it is resumed by an anaphoric pronoun. Example (19) shows the two options the informants were given, one with *niște* and one without:

- (19) Modern Romanian
 Am văzut (niște) tineri în fața clădirii. Numai
 have.PRS.1SG see.PTCP NIȘTE teenager.PL in face.ART building.GEN only
 doi dintre ei m-au salutat.
 two of.between they me.ACC-have.PRS.3PL greet.PTCP
 ‘I saw some teenagers in front of the building. Only two of them greeted me.’

The numbers are again very clear. In the overwhelming majority of cases, the speakers opted for the version containing *niște*, as Table 3 shows.

Table 3: Preferences with respect to presence/absence of *niște* with cataphorical DPs

Bare noun	<i>niște</i>
2 (6%)	30 (94%)

4.3.2 Translation task

The results of the translation task concerning the use of *niște* with specific nominals stem from two different sentences in the questionnaire. In one sentence, the respective DP is in direct object position (20), in the other the DP is the agent adjunct of the passivized verb *einladen* ‘to invite’ (21). Note that the DP is in both cases epistemically specific, in that the person uttering the respective sentence knows the persons the respective nominals denote:

- (20) German
 Im Restaurant habe ich Nachbarn getroffen, die du auch
 in.ART restaurant have.PRS.1SG I neighbor.PL meet.PTCP REL you also
 kennst: Paul und Erich.
 know.PRS.2SG Paul and Eric
 ‘In the restaurant I met neighbors you know, too: Paul and Eric.’

(21) German

Gestern wurde ich von Freunden eingeladen, die ich letztes
yesterday become.PST.1SG I by friend.PL invite.PTCP REL I last
Jahr kennengelernt habe, Lena und Marc.
year get.to.know.PTCP have.PRS.1SG Lena and Marc
‘Yesterday I was invited by friends whom I got to know last year, Lena
and Marc.’

The numbers in Table 4 show not only that about half of the informants opt for a definite article in the Romanian translation, but also that the use of a bare noun is still more frequent than the use of *niște*.

Table 4: Translations of epistemically specific indefinite German DPs

Syntactic function	Definite article	Bare noun	<i>niște</i>	Numeral	Demonstrative
Direct object	15 (48%)	9 (29%)	4 (13%)	2 (7%)	1 (3%)
Agent complement	15 (50%)	8 (27%)	6 (20%)	1 (3%)	–

4.4 Scope properties

The results concerning the scope properties of *niște* are subdivided in one subsection dealing with the scope properties of *niște* with respect to negation (§4.4.1) and one focusing on contexts where *niște* interacts with quantifiers (§4.4.2). The results of this whole section only stem from interpretation tasks.

4.4.1 Scope properties with respect to negation

As for the scopal behavior of *niște* in the context of negation, we have to distinguish between the results concerning *niște* in combination with a mass noun (22) and the ones dealing with *niște* preceding a plural count noun (23):

(22) Modern Romanian

N-am băut niște vin.
NEG-have.PRS.1SG drink.PTCP NIȘTE wine
‘I didn’t drink (any) wine.’

(23) Modern Romanian

N-am văzut niște tineri.
 NEG-have.PRS.1SG see.PTCP NIȘTE teenager.PL
 ‘I didn’t see (any) teenagers.’

In both cases the speakers were given two possible continuations of the respective sentence, one corresponding to a narrow-scope interpretation of *niște* (and thus putting the DP on a par with the corresponding bare noun), the other one to a reading where *niște* takes wide scope over the negation. In other words, in the latter interpretation, there was some (kind of) wine that was not drunk and some teenagers who were not seen, respectively.

The figures in Table 5 show rather clearly that the former reading is preferred both with mass and plural count nouns, but that – at least for some speakers – *niște* can also take wide scope with respect to negation.

Table 5: Scope properties of *niște* with respect to negation (mass vs. count nouns)

Type of noun	Narrow scope	Wide scope
Mass noun	27 (84%)	5 (16%)
(Plural) count noun	25 (78%)	7 (22%)

4.4.2 Scope properties with respect to quantifiers

The results concerning the scopal behavior of *niște* in the context of a quantifier stem from an interpretation task focusing exclusively on plural count nouns. As in the task described in the previous subsection, the speakers were given two sentences and two different readings from which they had to choose the one they preferred. As the position of the subject plays a crucial role in information structure in Romanian (cf. Leonetti 2017: 902) and might thus distort the findings, both postverbal (24) and preverbal (25) subjects have been included in the test:

(24) Modern Romanian

În fiecare duminică vin niște prieteni să ne
 in every Sunday come.PRS.3PL NIȘTE friend.PL COMP US.ACC
 viziteze.
 visit.PRS.SBJV.3PL
 ‘Every Sunday some friends come to visit us.’

(25) Modern Romanian

Niște copii vin să se joace aici în fiecare
 NIȘTE kid.PL come.PRS.3PL COMP REFL play.PRS.SBJV.3PL here in every
 zi.
 day
 ‘Some kids come to play here every day.’

Unlike in the task focusing on the scopal behavior with respect to negation, the two options contained the two different possible interpretations the sentence can have, rather than possible continuations. One interpretation corresponded to a reading where *niște* takes narrow scope over the quantifier and the other one to a wide-scope behavior of *niște*. The results differ clearly from the results concerning the scope properties of *niște* in the context of negation, in that *niște* does not seem to show any scope preferences at all – neither with a preverbal nor with a postverbal subject.¹¹ This is illustrated in Table 6, which contains the result for the reactions to (24) and (25).

Table 6: Scope properties of *niște* with respect to quantifiers

Subject position	Narrow scope	Wide scope
preverbal	16 (50%)	16 (50%)
postverbal	17 (53%)	15 (47%)

4.5 Combination with the DOM-marker *pe*

As far as the possible combination of *niște* with the DOM-marker *pe* is concerned, we have results from two different tasks: §4.5.1 presents the one from the preference task, §4.5.2 the one from the grammaticality judgment task. Again, the findings resulting from the different tasks differ remarkably from one another.

4.5.1 Preference task

There are, at least in principle, three conceivable possibilities when it comes to direct objects referring to a human being, depending on its degree of specificity:

¹¹As a reviewer pointed out, it might also be possible that the participants just weren’t sensitive to these distinctions.

(i) One version with the DOM marker *pe* but no additional clitic (which is, according to the literature, only marginally accepted, cf. §2.4); (ii) one with both clitic doubling and *pe*; and (iii) one with neither of them. The results of the preference task stem from one sentence in the questionnaire presenting these different versions, of which the informants had to indicate the one they preferred. Example (26) subsumes the versions (i) and (ii), (27) shows version (iii):

(26) Modern Romanian

Ieri (i-)a văzut pe niște studenți în bibliotecă.
 yesterday CL-have.PRS.3SG see.PTCP *pe* NIȘTE student.PL in library
 ‘Yesterday he saw some students in the library.’

(27) Modern Romanian

Ieri a văzut niște studenți în bibliotecă.
 yesterday have.PRS.3SG see.PTCP NIȘTE student.PL in library
 ‘Yesterday he saw some students in the library.’

The results in Table 7 show a clear preference for the absence of *pe* (iii) in combination with *niște*. However, version (ii) with *pe* and clitic doubling seems acceptable as well. Interestingly, two speakers even prefer version (i).

Table 7: Preferences with respect to the presence/absence of *pe* and CL in combination with *niște*

(i) [+ <i>pe</i>]	(ii) [+CL, + <i>pe</i>]	(iii) [- <i>pe</i>]
2 (6%)	7 (20%)	25 (74%)

4.5.2 Grammaticality judgment task

The results stemming from the grammaticality judgment task show a somewhat different picture insofar as the combination of *niște* and the DOM-marker *pe* and an additional clitic is not only regarded as “grammatical, but uncommon”, but even as “unproblematic” by the majority of the speakers. Consider the sentence (28) and its judgments (Table 8):

(28) Modern Romanian

Ieri i-a văzut pe niște nepoți de-ai
 Yesterday CL-have.PRS.3SG see.PTCP *pe* NIȘTE grandchild.PL of-AGR
 săi în bibliotecă.
 POSS.MPL in library
 ‘Yesterday he saw some of his grandchildren in the library.’

2 Romanian *niște* between non-specific and specific interpretations

Table 8: Grammaticality judgment concerning the combination of *niște* with *pe* and a clitic (direct object *in situ*)

Incomprehensible	Impossible	Possible, but unusual	Fully unproblematic
1 (3%)	1 (3%)	4 (13%)	26 (81%)

The usualness of the combination seems, however, to depend at least partially also on the syntactic function. If the noun phrase is in object predicative complement position (of verbs such as ‘to consider as’), the combination is considered equally grammatical, but much more unusual (Table 9).

Table 9: Grammaticality judgment concerning the combination of *niște* with *pe* and a clitic (predicative complement *in situ*)

Incomprehensible	Impossible	Possible, but unusual	Fully unproblematic
–	1 (3%)	13 (41%)	18 (56%)

A similar effect can be observed in the case of clitic left-dislocation, as in sentence (29).

(29) Modern Romanian

Mi-a zis că pe niște copii îi cunoaște
me.DAT-have.PRS.3SG tell.PTCP COMP *pe* NIȘTE child.PL CL know.PRS.3SG
de foarte mult timp.
of very much time
‘He told me that he’d known some kids for a very long time.’

Consider Table 10, containing the results of the judgments for (29) and two other sentences with a clitic left-dislocated noun phrase preceded by *pe* and *niște*.

Table 10: Grammaticality judgment concerning the combination of *niște* with *pe* (clitic left-dislocated direct object)

Incomprehensible	Impossible	Possible, but unusual	Fully unproblematic
1 (1%)	14 (15%)	23 (24%)	57 (60%)

Finally, when the noun phrase is clitic left-dislocated and its referent explicitly contrasted with another referent, average judgments of grammaticality sink significantly. Consider sentence (30) and Table 11, showing the results of the respective judgments given by the speakers:

- (30) Modern Romanian
 Pe niște copii i-am văzut, restul clasei
pe NIȘTE kid.PL CL-have.PRS.1SG see.PTCP rest.ART class.ART.GEN
 era deja plecată.
 be.PST.3SG already leave.PTCP.FSG
 ‘I saw some kids; the rest of the class had already left.’

Table 11: Grammaticality judgment concerning the combination of *niște* with *pe* (clitic left dislocated contrasted direct object)

Incomprehensible	Impossible	Possible, but unusual	Fully unproblematic
–	9 (30%)	7 (23%)	14 (47%)

5 Discussion

The aim of this section is to discuss the results presented in the previous chapter in some further detail. §5.1 is primarily concerned with the discussion of general findings, such as the non-obligatoriness – or, put differently, the frequent preference of a bare noun over a noun preceded by *niște* – and the apparent impossibility of using *niște* with generic nominals. However, it also tries to shed light on the general semantics of *niște* by discussing data stemming from introspection, i.e. the comments which the speakers were asked to make in the questionnaire. In §5.2 we turn our attention to the actual core topic of this chapter: the specificity-related properties of *niște*. We discuss the points supporting an analysis of *niște* as a specificity marker and the counterarguments some of our results represent. §5.3 discusses the scope properties of *niște* and compares them with the scope properties reported for the French and Italian PA.

5.1 Generalities

The first general and important observation emerging from the results of the translation task is the fact that *niște* is always optional, regardless of the syntactic

function of the DP of which it is a part.¹² It is only rarely used in active production and seems to be subject to individual preferences. In this respect, *niște* resembles the PA of Standard Italian, the use of which is said to be optional as well, at least from a purely syntactic viewpoint.

Let us now have a look at the sentences containing a plural count noun, which were translated at least three times by making use of *niște*.¹³

The six input sentences in (31–36) induced 23 occurrences of *niște*, which is more than half of all the occurrences found in the translation task. For reasons of convenience, the two sentences already given in (20) and (21) are repeated here as (31) and (32).

(31) German

Im Restaurant habe ich Nachbarn getroffen, die du auch
in.ART restaurant have.PRS.1SG I neighbor.PL meet.PTCP REL you also
kennst: Paul und Erich.
know.PRS.2SG Paul and Eric
'In the restaurant I met neighbors you know too: Paul and Eric.'

(32) German

Gestern wurde ich von Freunden eingeladen, die ich letztes
yesterday become.PST.1SG I by friend.PL invite.PTCP REL I last
Jahr kennengelernt habe, Lena und Marc.
year get.to.know.PTCP have.PRS.1SG Lena and Marc
'Yesterday I was invited by friends whom I got to know last year, Lena
and Marc.'

(33) German

Da waren Kinder im Laden, die ihre Mutter suchten.
there be.PST.3PL child.PL in.ART store REL their mother search.PST.3PL
'There were children in the store who were looking for their mother.'

(34) German

Äpfel hätte ich auch noch gerne.
apple.PL have.COND.1SG I also still please
'I'd also like to have some apples.'

¹²For the sake of completeness, it has to be mentioned that there is, in fact, one context where the use of *niște* seems obligatory, namely in combination with the comparative adverbial *ca* 'like'. As a consequence, *ca niște stăpâni* 'like (some) rulers' is different from *ca stăpâni* 'as rulers' (cf. Avram 1986: 82). This context is, however, not part of our study.

¹³The other two sentences which generated three or more translations showing *niște* contain a quantifier preceding a mass noun. In fact, the DP *etwas Kürbis* 'some squash' produced the highest number of occurrences of *niște* (12 occurrences). As this paper is more concerned with specificity-related properties of *niște*, we will not discuss this data any further here.

(35) German

Auf diesem Teller gibt es Eier.
 on DEM.MASC.SG plate give.PRS.3SG it egg.PL
 ‘There are eggs on this plate.’

(36) German

Es sind nur ein paar wenige Krümel übriggeblieben im
 it be.PRS.3PL only a few little.PL crumb.PL leave.OVER.PTCP in.ART
 Teller.
 plate
 ‘There are only some few crumbs left in the plate.’

As has already been mentioned in §4.3.2, in the translations of (31) and (32), *niște* competes not only with a bare noun, but to an even bigger extent with the definite article. The sentences (31) and (32) produce a total of 10 occurrences of *niște* (in 63 valid translations, which equals 16%). In the sentences (33) to (35), *niște* is predominantly in competition with a bare noun: There are 10 cases of *niște* and 62 occurrences of bare nouns.¹⁴ As for sentence (36), there is more variation: *niște* (3 occurrences) is considered an alternative to the quantifier *câteva* ‘some’, which clearly dominates in the translations (22 occurrences).¹⁵

We already mentioned that five persons used *niște* particularly often in their translations. By zooming in on these five speakers (abbreviated by “Sp.”), we can easily illustrate that the use of *niște* is strongly connected to individual preferences (Table 12).

The figures show that (i) 17 of the 23 occurrences generated by these six sentences stem from these five speakers and that (ii) the ratio of the use of *niște* in (31) to (36) is considerably higher among these speakers (47–70% vs. 10–16%). In these contexts, the use of *niște* is apparently a valid or even the preferred option for these five speakers.¹⁶

The obvious question which now arises regards the nature of the semantic difference between a bare noun and a noun preceded by *niște*, i.e. the question what *niște* contributes semantically to the meaning of the respective DP. In order

¹⁴ Additionally, there is one use of *unii* ‘certain’ and *cățiva* ‘some’, respectively, for (33), and one use of *câteva* ‘some’ for (34). The rate of occurrence of *niște* in the translations of these three sentences is thus 13% (10 out of 75).

¹⁵ In addition to *niște*, there are also three occurrences of *puține* ‘few’ as well as one use of a bare noun and *ceva* ‘some’ respectively. This amounts to a frequency of 10% with which *niște* is used in the translations (3 out of 30).

¹⁶ One might object that the presence/absence of *niște* in (33) could be caused by a difference between a distributive/collective reading and has nothing to do with individual preferences. Indeed, *niște* seems to strongly favor a collective reading: a distributive interpretation is, however, not excluded (contra Nedelcu 2009: 208; cf. Davatz 2018 for further details).

2 Romanian *niște* between non-specific and specific interpretations

Table 12: Distribution of the use of *niște* over the sentences (31) to (36) among the five speakers using it most frequently

	(31)	(32)	(33)	(34)	(35)	(36)
Sp. 1	✓	✓	✓	✓	✓	
Sp. 2		✓	✓	✓		✓
Sp. 3		✓				✓
Sp. 4	✓	✓	✓		✓	
Sp. 5		✓				✓
	7/10=70%		7/15=47%			3/5=60%

to answer this question and analyze the semantics of *niște*_{PL}, we shall look now at the comments made by the speakers in the preference task. Consider again the sentence given in (18), repeated below as (37):

(37) Modern Romanian

Mama a întâlnit (niște) membri ai
 mum.ART have.prs.3SG meet.PTCP NIȘTE member.PL GEN
 parlamentului: Șerban și Ioan.
 parliament.ART.GEN Șerban and Ioan

‘Mum has met (some) members of the parliament: șerban and Ioan.’

As was illustrated in §4.3.1, the speakers showed a clear preference for the version where *niște* precedes the specific direct object. However, the reasons for why they choose one or the other version varied considerably between the informants: (i) Three people explained their preference for *niște* by the (implicit) marking of a (low) quantity in the example, (ii) three other people explained it by the referential specificity of the direct object and (iii) one person motivated her preference for the non-use of *niște* by the fact that the referents of the noun *membri* ‘members’ are known, i.e. that they are epistemically specific. Consider the respective statements in 1–3:

1. Numind 2 oameni (Șerban și Ioan), avem nevoie de o marcă a cantității.
 ‘Naming two people (S. and I.), we need a marker of the quantity.’
2. Membrii sunt specificați, deci *niște* se potrivește.
 ‘The members are specified, so NIȘTE fits.’
3. Wir wissen schon, welche.
 ‘We already know who [it is about].’

Another general observation which can be made is that *niște_{PL}* is not actively used with generic nominals (cf. §4.2). This finding, resulting from the translation task, supports the claims made by Avram (1986: 82), stating that *niște* cannot have a “generic value”, which contradicts Nedelcu (2009: 207).

Niște_{PL} is thus distinct from the French plural PA, the use of which is possible with contrastive generic preverbal subjects (cf. Vogeleer & Tasmowski 2005: 69 and Wilmet 2003: 165):

(38) French

Des moutons n'ont jamais cinq pattes!

PA sheep.PL NEG-have.PRS.3PL never five paw.PL

‘Sheep *never* have five legs!’

However, there are in fact other Romance varieties showing PAs which behave similarly to *niște_{PL}*. Recent fieldwork in the Aosta Valley reveals that PAs with preverbal generic nominals are systematically translated by a definite article and never produced actively in the local Franco-Provençal varieties (see Stark & Gerards 2020, Ihsane 2018).¹⁷

The complete absence of *niște_{PL}* with generic nominals in active production does not, however, necessarily imply an actual ungrammaticality of this use. It remains to be tested whether and to what extent it is considered grammatical.¹⁸

5.2 *Niște_{PL}* as a specificity marker?

The question that has to be asked now is: Could the above-discussed absence (or even impossibility) of *niște* preceding a generic nominal in subject position be due to properties of specificity which are inherent to it? Providing an answer to this question using the collected data is, however, far from easy, as the results concerning the extent to which *niște_{PL}* can (or has to) be considered a marker of specificity differ remarkably depending on the task. When the speakers have the choice between a bare noun and a DP introduced by *niște*, they clearly opt for the latter, whereas when they are given a German bare noun, they tend to prefer a translation with a bare noun. The fact that German indefinite plural count nouns are undetermined might have an important influence on the translation; a comparison with translations of French sentences with indefinite plural count nouns determined by the so-called “partitive article” would allow us to further

¹⁷As for the use of Italian partitive articles in these contexts, see Cardinaletti & Giusti (2016: 77).

¹⁸Recent fieldwork conducted by the authors reveals that, though categorically avoided in active production, generic PA-subjects are not considered ungrammatical by all the speakers in grammaticality judgments (see Davatz et al. 2023).

explore this possible factor. Another conceivable explanation for the difference between the results of the two tasks is the educational background of the participants: Even though asked to translate the sentences as naturally as possible into Romanian, many of them might have aimed at a stylistically high rather than a “spontaneous” and “natural” translation.

The preference task shows that the speakers clearly prefer (88%) the use of *niște_{PL}* over a bare noun if the respective noun phrase in direct object position is epistemically specific (cf. §4.3.1). As shown in the previous section, the reasons why they do so seem to differ, but one reason mentioned by various participants was the “specificity” of the noun phrase. Other speakers assign their choice for *niște_{PL}* to the given (and highly restricted) number of members of the parliament, which are denoted by the noun phrase in the respective example (cf. 37). This implies that *niște_{PL}* is preferentially used to denote smaller quantities, whereas a bare noun is not specified at all for the quantity of referents it denotes. Three further comments made by the speakers seem to support this observation, the first one in 4 related to (37), 5 and 6 to (19):

4. Mama a întâlnit câțiva membri, puțini.
‘The mother has met *some* members, *few*.’
5. *Niște* are rolul de a indica atât un număr redus de indivizi, cât și de a oferi specificitate complementului direct.
‘*Niște* has both the role of *indicating a reduced number of individuals* and of *conferring specificity* on the direct complement.’
6. Ohne die Angabe *niște* kann die Rede von 2 oder 20 Jugendlichen sein.
‘Without the indication *niște* it can be about 2 or 20 teenagers.’

Considering the numbers (Table 3) and comments for the sentence in (19), repeated under (39), which shows a very high rate of uses of *niște_{PL}* despite the non-epistemic specificity of the direct object, it seems that a small quantity might indeed be the more important factor than epistemic specificity when it comes to the use of *niște_{PL}*:

- (39) Modern Romanian
 Am văzut (niște) tineri în fața clădirii. Numai
 have.PRS.1SG see.PTCP NIȘTE teenager.PL in face.ART building.GEN only
 doi dintre ei m-au salutat.
 two of.between they me.ACC-have.PRS.3PL greet.PTCP
 ‘I saw some teenagers in front of the building. Only two of them greeted me.’

Judging from the comment in 5, the two notions of “small quantity” and “specificity” seem strongly intertwined for the speakers. It is, in our opinion, not excluded that the “specificity” effect is an implicature of *niște* referring to a small quantity: What is reduced in number implicates a higher degree of specificity, a higher probability of “referential anchoring” (see §2.3). Another conceivable factor for the preferred use of *niște_{PL}* in (39) with the anaphoric pronoun *ei* is one that has been brought up for the DOM-marker *pe* (cf. Chiriacescu & von Heusinger 2009, 2010; von Heusinger & Chiriacescu 2013), namely discourse prominence. It might be that marking a DP by *niște_{PL}* increases “the potential to generate further co-referential expressions”, as Chiriacescu & von Heusinger (2009: 13) state it for *pe*.

The results of the translation task (cf. §4.3.2) show that *niște_{PL}* is often omitted in similar contexts and used less frequently than a bare noun (Table 4). This suggests that epistemic specificity might be more a result of the presence of *niște_{PL}* than a (strong) trigger for its use, similarly to what Klein & de Swart (2011) stated for DOM-markers (and confirming the findings of Dobrovie-Sorin 2013).

Interestingly, however, the combination with the DOM-marker *pe* marking scopal and epistemic specificity is clearly dispreferred in the preference task (cf. §4.5.1.). Yet this combination is not considered ungrammatical, but seems to be restricted to very specific cases where *niște_{PL}* is interpreted partitively.¹⁹

5.3 *Niște_{SG/PL}* with surprising scope properties

Beyond the insights presented in the two previous subsections and the apparent difficulties in pinning down the exact semantics of *niște*, our pilot study additionally shows that *niște* differs from other indefinite determiners like the PA in French and Italian with respect to (some of) its scope properties.

The results of the interpretation task, designed specifically to reveal the scopal behavior of *niște* show (i) an apparent scopal indifference with respect to other quantifiers for *niște_{PL}*, and (ii) a clear domination of narrow scope of *niște_{SG/PL}* with respect to negation, with, however, *no systematic exclusion of wide scope* (cf. §4.4).

As far as (i) is concerned, the literature shows that the same holds true for the French PA (cf. Ihsane 2008: 139). The second observation is, however, much more surprising, in that *niște* seems, at least for some speakers, to be able to scope over the negating element both with plural count nouns *and mass nouns*.

¹⁹Regarding the acceptability of the sentence *Îi consideră inteligenți pe niște copii* ‘He considered some children intelligent’, one speaker commented that it would be acceptable *doar dacă DOAR pe unii dintre ei* ‘only if ONLY some of them’.

While the former is also true for the Italian PA (cf. Cardinaletti & Giusti 2016: 60), the latter is attested neither for the French nor the Italian singular PA (cf. Ihsane 2008: 139f. for French; Cardinaletti & Giusti 2016: 60 for Italian). And *niște*_{PL} is not systematically associated with wide scope, as one might expect from its apparent preference for specific DPs: It *can* be interpreted having wide scope, but does not have to be.

6 Conclusion

Coming back to the general research question building the background of this article, i.e. the question whether *niște* is semantically comparable to the so-called “partitive articles” of French and Italian, we can state similarities and differences. §4 and 5 have helped answer our three detailed research questions set up in §2.5. First, compared to bare plural count nouns, *niște*_{PL} seems to favour a specific interpretation, but is maybe not always compatible with the DOM-marker *pe* (plus clitic-doubling). Second, *niște*_{PL} is incompatible with generic readings/contexts, and third, the scope properties of *niște* are quite idiosyncratic.

Even if there are thus some characteristics which *niște* seems to share with the so-called “partitive articles” of French and Italian – like the impossibility of use with generic subjects or possible wide scope with respect to quantifiers in the plural – there are two crucial properties which clearly distinguish it from them. First and foremost, *niște*_{SG} is apparently able to scope over negation, whereas wide scope with respect to negation is unattested both for the French and Italian singular PA. The second property distinguishing *niște*_{PL} from the two other plural PAs is the fact that it seems to be used preferentially in the context of epistemic specificity or subsequent anaphoric pronouns. However, the compatibility with the direct object marker *pe*, reported to be a marker of specificity, is relatively low. The comments made by the speakers insinuate that the meaning of *niște*_{PL} is slightly different in this case: *Niște*_{PL} seems to denote a part of a whole. It seems conceivable that *niște*_{PL}, similarly to the DOM-marker *pe*, is a marker of specificity or, maybe even more to the point, of discourse prominence (cf. Chiriacescu & von Heusinger 2009, 2010, von Heusinger & Chiriacescu 2013), and that their co-occurrence thus leads to a clash due to redundancy. The mechanics causing the different interpretation of *niște* in this context are, however, yet to be understood, and an analysis of their precise interaction is called for in future research. What is clear already at this stage is that *niște*_{SG/PL} has to be considered as an element of its own and is only partially comparable to other Romance indefinite determiners.

References

- Avram, Mioara. 1986. *Gramatica pentru toți*. București: Editura Academiei Republicii Socialiste România.
- Cardinaletti, Anna & Giuliana Giusti. 2016. The syntax of the Italian indefinite determiner *dei*. *Lingua* 181. 58–80.
- Carlson, Greg N. 1977. *Reference to kinds in English*. University of Massachusetts. (Doctoral dissertation).
- Chiriacescu, Sofiana & Klaus von Heusinger. 2009. Pe-marking and referential persistence in Romanian. In Arndt Riester & Edgar Onea (eds.), *Working Papers of the SFB 732 "Incremental Specification in Context"*, 1–19. Online Publikationsverbund der Universität Stuttgart (OPUS).
- Chiriacescu, Sofiana & Klaus von Heusinger. 2010. Discourse prominence and pe-marking in Romanian. *International Review of Pragmatics* 2.2. 298–332.
- Cornips, Leonie & Cecilia Poletto. 2005. On standardising syntactic elicitation techniques. *Lingua* 115(1). 939–957.
- Davatz, Jan. 2018. *Der Indefinitmarker niște als Pendant zum französischen article partitif? Eine semantische und syntaktische Analyse*. University of Zurich. (MA thesis).
- Davatz, Jan, Tabea Ihsane & Elisabeth Stark. 2023. Enquêtes dialectologiques à Evolène: Les articles dits ‘partitifs’ et leurs équivalents. In Dorothée Aquino-Weber, Sara Cotelli Kureth, Andres Kristol & Aurélie Reusser-Elzingre (eds.), *“Coum’on étèila que kòoule... Come una stella cadente... Comme une étoile filante...” Mélanges à la mémoire de Federica Diémoz*, 65–85. Geneva: Librairie Droz.
- Davatz, Jan & Elisabeth Stark. 2021. Niște en roumain – quantifieur ou déterminant? In Lene Schøsler, Juhani Härmä & Jan Lindschouw (eds.), *Actes XXIXe Congrès International de Linguistique et de Philologie Romane de Copenhague, 1er–6 juillet 2019*, 401–414. Strasbourg: ELiPhi.
- Dobrovie-Sorin, Carmen. 2013. Bare nouns. In Carmen Dobrovie-Sorin & Ion Giurgea (eds.), *A reference grammar of Romanian*, vol. 1: The noun phrase (Linguistik Aktuell/Linguistics Today 207), 49–96. Amsterdam: John Benjamins.
- Farkas, Donka. 1995. Specificity and scope. In Léa Nash & George Tsoulas (eds.), *Actes du Premier Colloque Langues & Grammaire*, 119–137. Paris: Garland.
- Romalo, Valeria Guțu (ed.). 2005. *Gramatica Limbii Române: Cuvântul*, vol. 1. București: Editura Academiei Române.
- Hill, Virginia & Alexandru Mardale. 2017. On the interaction of differential object marking and clitic doubling in Romanian. *Revue Roumaine de Linguistique* 62(4). 393–409.

- Ihsane, Tabea. 2008. *The layered DP in French: Form and meaning of French indefinites*. Amsterdam: John Benjamins.
- Ihsane, Tabea. 2018. *Preverbal subjects with a partitive article: The case of Franco-provençal in the Aosta Valley*. Societas Linguistica Europaea (SLE 2018). University of Tallinn, Estonia.
- Klein, Uli & Peter de Swart. 2011. Case and referential properties. *Lingua* 121. 3–19.
- Leonetti, Manuel. 2017. Basic constituent orders. In Andreas Dufter & Elisabeth Stark (eds.), *Manual of Romance morphosyntax and syntax*, 887–932. Berlin: De Gruyter.
- Milsark, Gary L. 1974. *Existential sentences in English*. MIT. (Doctoral dissertation).
- Nedelcu, Isabela. 2003. ‚Niscai‘ observații despre ‚niște‘. In Gabriela Pană Dindelegan (ed.), *Aspecte ale dinamicii limbii române actuale. Actele colocviului Catedrei de Limba Română, II, Section Morfosintaxă, 27-28 noiembrie 2002*. București: Editura Universității din București.
- Nedelcu, Isabela. 2009. *Categoria partitivului în limba română*. București: Editura Universității din București.
- Pană Dindelegan, Gabriela. 2016. *The syntax of Old Romanian*. Oxford: Oxford University Press.
- Stan, Camelia. 2006. ‚Articolul‘ niște. In Gabriela Pană Dindelegan (ed.), *Limba Română: Stadiul Actual al Cercetării, actele celui de al 6-lea colocviu al catedrei de limba română*, 199–205. București: Editura Universității din București.
- Stark, Elisabeth. 2011. Fonction et développement du marquage différentiel de l’objet direct (MDO) en roumain, en comparaison avec l’espagnol péninsulaire. In Société de Linguistique de Paris (ed.), *Mémoires de la Société de Linguistique de Paris 19: L’évolution grammaticale à travers les langues romanes*, 35–61. Leuven: Peeters.
- Stark, Elisabeth & David Gerards. 2020. “Partitive articles” in Aosta Valley Francoprovençal: Old questions and new data. In Tabea Ihsane (ed.), *Disentangling bare nouns and nominals introduced by a partitive article*, 301–334. Leiden: Brill.
- Vogeleer, Svetlana & Liliane Tasmowski. 2005. Les N, Un N et Des N en lecture générique. *Travaux de Linguistique* 50. 53–78.
- von Heusinger, Klaus. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics* 19. 245–274.
- von Heusinger, Klaus & Sofiana Chiriacescu. 2013. The discourse structuring potential of differential object marking, the case of indefinite and definite direct objects in Romanian. *Revue roumaine de linguistique* 58.4. 439–456.
- Wilmet, Marc. 2003. *Grammaire critique du français*. Brussels: Duculot.

Chapter 3

Argentinian Spanish *cualunque* and Italian *qualunque*

Marika Francia

Georg-August-Universität Göttingen

Olga Kellert

Georg-August-Universität Göttingen

In this article, we describe the syntactic and semantic properties of Argentinian Spanish *cualunque* ‘common, unremarkable’, which originated from Italian *qualunque* ‘any’. As we will show, on the basis of different corpus data (mainly social media data), *cualunque* has a syntactic distribution different from that of Italian *qualunque*.

Our main hypothesis is that Arg.Sp. *cualunque* denotes a property of type $\langle e, t \rangle$, such as *es cualunque*, and has the meaning of ‘unremarkable’. It describes a property of atomic individuals (e.g., people or objects) or their kinds (see Carlson 1977) that do not have any particular or specific features that would distinguish them from other atomic individuals or their kinds. Thus, *cualunque* entails a contrast to specific or particular (kinds of) individuals with special and distinguished properties. The crucial point of the analysis suggested in this paper is that the atomic individuals or their kinds described by *cualunque* are evaluated on different scales, such as a frequency scale or some qualitative scale (e.g., scale of goodness). The neutral interpretation of *cualunque* as ‘common’, ‘widespread’, or ‘normal’ is yielded when the individuals described by *cualunque* are evaluated in the middle of these scales. The pejorative version of the meaning ‘common’ is analyzed as a predication over individuals or their kinds that are evaluated on the extreme end of a qualitative scale, such as the scale of goodness. On this interpretation, *cualunque* means ‘not outstanding’ or ‘bad’. It is the evaluation on scales that has triggered the reanalysis of *cualunque* as a gradable adjective or as a noun with the evaluative interpretation of ‘ordinary or bad person’ in Argentinian Spanish. Italian *qualunque* also expresses the evaluation on scales, but only as a pragmatic implicature, which is not (yet) lexicalized.



1 Introduction

Italian had an invasive influence on the variety of Spanish spoken in the Río de la Plata area. Its effects are visible in all aspects of the language, especially in the lexicon, which has been enriched by many Italianisms (see Meo Zilio 2001, Di Tullio 2003, Engels & Kailuweit 2011, among others). Argentinian Spanish (Arg.Sp.) *cualunque* is one of them.

The influence of the Italian language and its dialects on Argentinian Spanish was the result of prolonged and intense linguistic contact between the local population and the Italian community that had emigrated to Argentina in the 19th through mid-20th century (see Meo Zilio 1959, Di Tullio 2003, Conde 2011: pt. 2.1).¹ Most of these immigrants were illiterate, spoke only their local dialect, and had only a passive knowledge of Standard Italian (see De Mauro 2017 among others). In Argentina, they settled in urban environments, such as the city of Buenos Aires (see Baily 1999: pt. 2, Devoto 2002: 41). Figure 1 offers an overview of this historical context. It shows two maps that represent the Italian regions most affected by the migratory flow and the areas of major settlement by Italian immigrants in Argentina. Among these, there is the city of Buenos Aires, where the argot known as Lunfardo² was born, and a region adjacent to that city, where its use spread in the first half of 20th century (see Teruggi 1974: 32, Conde 2011: pt. 1.2, among others).³

Arg.Sp. *cualunque* is one of the many Italianisms that enriched Rioplatense Spanish. Its origin can be traced back to Italian (It.) *qualunque* ‘any’ (see Real Academia Española 2010, Haensch & Werner 2000: 223) or to a similar-sounding dialectal form, such as Piedmontese *qualonque* or Sicilian *qualunchi* (Brero 2001: 213, Piccitto & Tropea 1990: 1041). The exact historical development of *q/cualunque* is difficult to reconstruct due to a lack of sources of spoken Italian

¹Considering only the years of mass migration (1881–1914), 2,000,000 Italians arrived in Argentina. In 1914 Italians constitute ca. 12% of the entire Argentinian population (see Devoto 2002: 32ff.).

²Lunfardo is an argot, about 40% of whose vocabulary is derived from Italian. Its origin goes back to Buenos Aires in the second half of 19th century. Its use later spread first throughout the entire region and subsequently across the whole country, penetrating the spoken language of the Spanish-speaking population (see Teruggi 1974, Conde 2011).

³The map on the right in Figure 1 shows the Italian regions with the heaviest emigration from 1876 to 1915 as reported by the Italian Commissariat of Emigration. The map on the left in Figure 1 shows the regions of settlement in Argentina according to the Third Census of the Argentinian Republic in 1914.

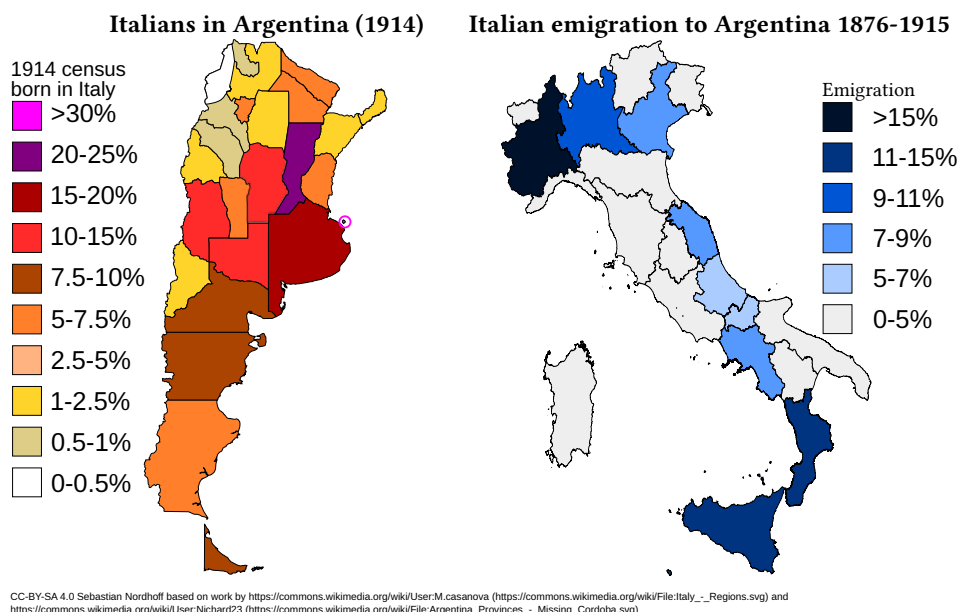


Figure 1: Regions of emigration from Italy and regions of Italian settlement in Argentina. Left: Adaptation of Italians in Argentina (1914), from [https://commons.wikimedia.org/wiki/File:Italians_in_Argentina_\(1914\).png](https://commons.wikimedia.org/wiki/File:Italians_in_Argentina_(1914).png). Right: Adaptation of Map 1 from Bailly (1999: 615).

and Spanish from that time. The first available diachronic source of *q/cualunque* starts with Cocoliche⁴, a *learner variety* of Spanish spoken by Italian immigrants and its imitation in literary contexts, so-called literary Cocoliche.⁵ The word was

⁴Cocoliche refers to the variety of Spanish that was spoken by Italian immigrants of first generation in Argentina in their everyday communication in the late 19th and early 20th centuries. It was the result of language contact with colloquial Spanish spoken in the Río de la Plata area and Italian dialects at that time. It disappeared with the second generation. Cocoliche is characterized in the literature as an individual, spontaneous, and unconscious phenomenon with high instability and irregularity, as it could differ between speakers according to many factors, e.g., language of origin, level of education, permanence of residence in the country, and willingness to learn (see Cancelleri 2001: 73f. Di Tullio 2003: 51f. Engels 2010: 44f. Kailuweit 2004: 54, Meo Zilio 1964).

⁵In accordance with Kailuweit (2004), we refer to the imitation of Cocoliche in literary contexts, such as plays, novels, and short stories, as literary Cocoliche. Kailuweit (2004: 59) and Ennis (2015: 129f.) define this literary variety as a form of conceptual orality in the sense of Koch & Oesterreicher (2011). It is based on the perception that each author has with respect to the Spanish spoken by Italian immigrants. As a result, it can vary from one author to another with respect to linguistic elements they consider characteristic. At the same time, these linguistic elements form a literary tradition from which subsequent authors can benefit (Kailuweit 2004: 66).

later attested in Lunfardo (see Conde 2011) and then in an informal style of speech in Argentinian Spanish, in the sense of Koch & Oesterreicher (2011).⁶

Arg. Sp. *cualunque* has changed phonologically, semantically, and syntactically from the original source word It. *qualunque*.

In Italian, *qualunque* is an indefinite, which can be used as a prenominal determiner with a Free Choice (FC) interpretation in modal contexts. Thus, when embedded under an overt modal as in (1), it is interpreted as ‘every possibility is an option’ (see Chierchia 2006, Aloni & Port 2013, Alonso-Ovalle & Menéndez-Benito 2017, Stark 2006, Kellert 2021b, among others).

(1) Italian (Kellert 2021b: 2)

Puoi scegliere **qualunque** **libro**
can.PRS.2SG choose QUALUNQUE book⁷

‘You can choose any book’

Conventional meaning: ‘You can choose a book’ and FC meaning: ‘each book is a possible option’.

Italian *qualunque* is also used as a postnominal indefinite with the meaning ‘ordinary/unremarkable’, often with copular verbs in the indicative present or past tense as in (2) and with indefinite nouns that assert the existence of some entity (see Kellert 2021c for Italian and Alonso-Ovalle & Menéndez-Benito 2017 for Spanish *cualquiera*). When postnominal *qualunque* is used in a predicate position as in (2), it adds an evaluative interpretation, namely that ‘the book is unremarkable’.

(2) Italian (Kellert 2021b: 2)

The Stars è un libro **qualunque**

The Stars be.PRS.3SG a book QUALUNQUE

The Stars is a book (existential inference)

The Stars is unremarkable (evaluative meaning)

⁶We have found the first diachronic occurrences in texts of *cocolichesco* characters in plays and novels dating back to the beginning of the 20th century (see Section 7). Following the ethnolectal chain described by Engels & Kailuweit (2011: 238), we assume that the term was first used by Italians speaking Cocoliche in their everyday attempts to communicate with the hispanophone population (*first variety*). At the same time, it could be heard and read in popular media, such as in *sainetes*, tangos, and popular magazines (*secondary ethnolect*). The popularity of *cocolichesco* characters led to its ironic imitation by native Spanish-speakers outside the context of the media (*tertiary ethnolect*). In this way, *cualunque* entered as an Italianism into the Lunfardo lexicon and later into the Argentinian Spanish informal style of speech.

⁷Since It. *qualunque* and Arg.Sp. *cualunque* have a number of different interpretations, we will not translate them in the gloss.

Even though postnominal *qualunque* is possible in non-modal contexts such as with indicative present or past tense copular verbs as in (2) – something unpredicted by standard analyses of Free Choice Indefinites such as Chierchia (2006) – *qualunque* seems to still have the status of an indefinite rather than that of an adjective, as most standard syntactic tests targeting the category of adjectives do not apply to *qualunque* in Standard Italian (see Kellert 2021b). For instance, It. *qualunque* cannot be used as a predicate under copular verbs as in (3a), in comparative structures (4a,b), with degree modification (5), in coordination with other adjectives (6), and in uses with definite nouns (7) or quantifiers (8).⁸

(3) Predicative *qualunque* (Italian)

- a. * Gianni è qualunque
 Gianni be.PRS.3SG QUALUNQUE
 ‘Gianni is qualunque’
- b. Gianni è un uomo qualunque
 Gianni be.PRS.3SG a man QUALUNQUE
 ‘Gianni is an ordinary man’

(4) Comparatives (Italian)

- a. * Gianni è più/meno/tanto qualunque di/quanto te
 Gianni be.PRS.3SG more/less/as QUALUNQUE than/as you
 ‘Gianni is more/less/as qualunque than/as you’
- b. * Gianni è tanto qualunque quanto te
 Gianni be.PRS.3SG as QUALUNQUE as you
 ‘Gianni is as qualunque as you’

(5) Degree modification (Italian)

- Gianni è un uomo (*molto/*abbastanza/*poco) qualunque
 Gianni be.PRS.3SG a man very/pretty/little QUALUNQUE
 ‘Gianni is a (*very/*pretty/*not) ordinary man’

(6) Coordination (Italian)

- Gianni è un uomo (*molto/*abbastanza/*poco) qualunque
 Gianni be.PRS.3SG a man very/pretty/little QUALUNQUE
 ‘Gianni is a (*very/*pretty/*not) ordinary man’

⁸A linguistic experiment conducted in 2020 on 160 native Italian speakers from different regions of Italy suggests that It. *qualunque* may be on its way to being reanalyzed as an adjective, as some informants accept uses of the word with coordination, degree modification, and comparatives (see Kellert 2021d for an analysis of an online survey conducted by Francia 2020). Further research needs to be conducted on It. *qualunque*.

- (7) Definites/Demonstratives (Italian)
 *questa/*la ragazza qualunque
 this/the girl QUALUNQUE
 ‘this/the girl qualunque’
- (8) Quantifiers (Italian)
- a. *tutte le/*tante/*due ragazze qualunque
 all/many/two girls QUALUNQUE
 ‘all/many/two girls qualunque’
 - b. *nessuna/*qualche/*ogni ragazza qualunque
 no/some/each girl QUALUNQUE
 ‘no/some/each girl qualunque’

Unlike It. *qualunque*, Argentinian Spanish *cualunque* is used in all aforementioned syntactic contexts in (3–8). This is shown by degree modification in (9), a comparative clause in (10), coordination (11), morphological agreement with the noun *cualunque* modified in (12), a definite determiner in (13), and a quantifier in (14). Our hypothesis is that *cualunque* can be used as an adjective with the meaning ‘ordinary/common/unremarkable/low-class/bad’ (Real Academia Española 2010, Haensch & Werner 2000: 223). It can also be used as a noun with a similar adjectival meaning of ‘an ordinary (male/female) person’ in (13).

- (9) Degree modification (Argentinian Spanish)
 es un nombre re cualquier
 be.PRS.3SG a name very CUALUNQUE
 ‘It’s a very ordinary name’
 (Twitter, 2019, <https://twitter.com/lauritula/status/1103348577324154880>)
- (10) Comparatives (Argentinian Spanish)
 Más cualquier que gato atigrado
 more CUALUNQUE than cat tabby
 ‘It’s more common than a tabby cat’
 (Twitter, 2019,
https://twitter.com/Felicitas73_/status/1126985847314767872)
- (11) Coordination (Argentinian Spanish)
 qué cartel [...] ordinario y cualquier
 what banner [...] ordinary and CUALUNQUE
 ‘what an ordinary and unremarkable banner’
 (Twitter, 2020, <https://twitter.com/sebyo53/status/660538375246626816>)

- (12) Agreement (Argentinian Spanish)
 A todos los periodistas [...] cualquieres
 to all.PL the.PL journalist.PL [...] CUALUNQUE.PL
 ‘to all ordinary journalists [...]’
 (Twitter, 2016, <https://twitter.com/anaopera/status/704785632200138752>)
- (13) Definites (Argentinian Spanish)
 el cualquier es el que dice que la culpa la
 the.M.SG CUALUNQUE.SG be.PRS.3SG the that say.PRS.3SG that the fault it
 tienen otros.
 have.PRS.3SG others
 ‘the ordinary man is the one who says that others are at fault.’
 (*Tiempo Argentino*, 2018,
<https://www.tiempoar.com.ar/politica/damian-selci-el-militante-es-el-producto-mas-civilizado-que-puede-tener-una-sociedad/>)
- (14) Quantifiers (Argentinian Spanish)
 Todo virus cualquiere
 all virus CUALUNQUE
 ‘all common viruses’
 (Twitter, 2016,
<https://twitter.com/InfoGeneral4p/status/772112368461545473>)

Arg.Sp. *cualunque* can also be used as an indefinite with a Free Choice interpretation in modal contexts, such as imperatives (15), which we interpret as a modalized construction: ‘You can put any shirt’. However, *cualunque* is less frequently used in modal contexts, as we will show using the frequency distribution in §4.

- (15) Ponete una remerita cualquier y una calza: ¡listo!
 put.IMP.2SG a shirt CUALUNQUE and a leggings ready
 ‘Put on any shirt and a pair of leggings: ready!’
 (Twitter, 2019,
<https://twitter.com/LaSuvuelabajo/status/1179938401056952320>)

To sum up, Argentinian Spanish *cualunque* can be used as a gradable adjective, a noun with a meaning similar to that of a gradable adjective, and as a Free Choice indefinite (noun modifier).

It is important to note that Argentinian Spanish has another indefinite element, namely *cualquiera* ‘any’. This word is normally used as a Free Choice pronoun in Arg.Sp. and in other Spanish varieties (e.g., *Cualquiera puede hacerlo*. ‘Anybody

can do this.’). However, in Argentinian Spanish, in contrast to European Spanish, it can also be used as an adjective, similar to *cualunque* (see Rizzo Salierno 2013, Kellert 2021a):

- (16) es re/tan/muy cualquiera
 be.PRS.3SG very CUALQUIERA
 ‘it’s very ordinary’
- (17) Un día re/tan/muy cualquiera
 a day very CUALQUIERA
 ‘a very ordinary day’

Rizzo Salierno (2013) assumes that *cualquiera* has changed into an adjective as a result of analogy to *cualunque*. However, *cualquiera* and *cualunque* do not have similar uses in any respect. *Cualquiera* in contrast to *cualunque* can also be used as an indefinite under transitive verbs such as *hacer* with the meaning ‘something bad’ and under *mandar* with the meaning ‘incorrect/false’ (Di Tullio 2015, Kellert 2021a).

- (18) (Kellert 2021a)
 Hizo cualquiera
 do.PST.3SG CUALQUIERA
 ‘She/He did something bad.’
- (19) (Kellert 2021a)
 Es así o mande cualquiera?
 be.PRS.3SG like.this or say CUALQUIERA
 ‘Is it true or is it false?’

Moreover, *cualunque* and *cualquiera* are diatopically different as the geolocation of the tweets with these lexical items have shown (see §2). The latter is used everywhere in Argentina, whereas *cualunque* is used in a more restricted geographical area (in big cities such as Buenos Aires and Cordoba, where Italian immigrants have settled). Moreover, *cualunque* is used by a particular social group that has some Italian background or identifies with this background as the user profile analysis as well as the content analysis of the tweets show (see Kellert 2021d for details).⁹ For this reason, due to their syntactic, semantic, and

⁹The relation between cultural background and use of *cualunque* needs to be tested in the future on a large number of quantitative results. All examples studied so far seem to show some relation to Italy or Italian products:

(i) nada que envidiar al Barilla, que son fideos cualesques en Italia.
 ‘measuring up to Barilla, which are ordinary noodles in Italy’
 (Twitter, 2015, <https://twitter.com/ElTanqueMartin/status/641040975561904129>)

sociolinguistic differences, we will refrain from pursuing a deeper analysis of *cualquiera* and *cualunque* in this article.

The main aim of this article is to understand the differences between modern Arg.Sp. *cualunque* and modern It. *qualunque* and to explain the conditions for variation and change with these lexical words.

This article is structured as follows: §2 provides a description of the methodology and corpora used for our investigation, as well as an overview of the graphical forms of Arg.Sp. *cualunque*. §3 describes diaphasic and diastratic factors of linguistic variation and correlates them to different orthographic forms found in the corpora. §4 describes the frequencies of syntactic properties of Arg.Sp. *cualunque*. §5 offers a semantic analysis of Arg.Sp. *cualunque*. §6 suggests hypotheses of the semantic and syntactic change in Argentinian Spanish. §7 gives a summary and an outlook for future research.

2 Data and methodology

In order to investigate the syntactic and semantic distribution of Arg.Sp. *cualunque*, we mainly based our research on corpus data from different sources. The data were collected from the *Corpus del Español* (CDE) by Mark Davies,¹⁰ in particular the CDE Web/Dialect and CDE News On the Web (CDE NOW). Both corpora are fully annotated and tagged for Part of Speech (PoS) and allow the user to limit the query to Argentinian Spanish. We also investigated data from a selected set of Google Books *n*-grams (BYU)¹¹ and from the online database *The Internet Archive*,¹² as well as data from Twitter localized in Argentina.¹³ We informally asked five native speakers of Argentinian Spanish about the use of *cualunque*. As to graphical variants, in order to cover a large range of forms,

¹⁰*Corpus del Español* by Mark Davies consists of five fully annotated and PoS tagged corpora: CDE (Genre/Historical); CDE Web/Dialect; CDE NOW (News On the Web); WordAndPhrase; Google Books *n*-grams (BYU). They are all accessible online via <https://www.corpusdelespanol.org>.

¹¹Google Books *n*-grams (BYU) is based on Google Books *n*-grams data. It consists of 45 billion words from 10 million books in Spanish published since the 19th century. It allows searching by part of speech, lemma, word, wildcards, synonym, syntactic construction, and collocations. However, it does not allow one to select a geographical region. We therefore selected a sample of books published in Argentina. The corpus is accessible online at <http://www.english-corpora.org/googlebooks/>.

¹²The digital library *The Internet Archive* provides approximately 250,000 books and texts in Spanish. The online database allows a search query by single terms and phrases, including Boolean operators. The data can be filtered by publication date and language as well. It is available at <https://archive.org>.

¹³The search on Twitter (<https://twitter.com>) was limited to tweets in Spanish published between 2006 and 2019. We considered 238 examples that were geotagged from Argentina.

we ran a search query for the sequence *alun* on CDE corpora, as well as for *cualunque*, *cualunques*, *cualunche*, *cualunches*, *qualunque*, *qualunques*, *qualunche*, *qualunches*, *qualunje*, *qualunjes*, *cualungue*, *cualunques*, *gualunque*, *gualunques*, *gualungue*, *gualungues*, *gualunche*, *gualunches* on *The Internet Archive* and on Twitter.

As a result, we obtained a corpus of 360 occurrences of *cualunque* in Argentinian Spanish, which includes the graphical forms listed in their order of frequency in Table 1.

Table 1: Graphical variants of Arg.Sp. *cualunque*

Graphical variants	<i>n</i>	%
<i>cualunque</i>	239	66.4%
<i>cualunche</i>	31	8.6%
<i>cualunques</i>	30	8.3%
<i>cualungue</i>	23	6.4%
<i>qualunque</i>	17	4.7%
<i>cualunches</i>	11	3.1%
<i>cualunje</i>	5	1.4%
<i>cualungues</i>	1	0.3%
<i>gualunque</i>	1	0.3%
<i>gualungue</i>	1	0.3%
<i>qualunques</i>	1	0.3%
Total	360	100%

The graphical variant *cualunque* is the most widespread form. The form *qualunque*, which is orthographically identical with It. *qualunque*, is less frequent than *cualunque*. The plural forms *cualunques* and *cualunches* are especially interesting, as Italian *qualunque* is invariant. The form *cualunche* with *-che(s)* is a mixed representation of Sp. *-cual-* and It. *-che* (lit. ‘that’), which might be explained in the future by the influence of Italian dialects spoken by Italian immigrants in Argentina or by an incorrect parsing of *-que* as the Italian complementizer *che* ‘that’.¹⁴ We will study these and other hypotheses that explain the variety of graphical forms in detail in the future.

¹⁴The Sicilian dialect has the form *qualunchi* (see Piccitto & Tropea 1990: 1041). In order to explain *cualunche* with the ending *-che*, we might assume that Sicilian orthography plays a role in the Arg.Sp. form *cualunche*. However, this assumption is problematic, given that it does not explain the phonological change of *-i* to *-e*. Another hypothesis is that speakers who use *cualunche* have parsed *-que* as the Italian complementizer *che* ‘that’, not knowing that *qualunque* is derived from Latin *unquam* and thus cannot be decomposed into *un-que*.

In the following section, we looked at whether different graphical forms correlated with specific register variations.

3 Correlation of form and register

The use of the most widespread form *cualunque* appears in contexts typically associated with an informal register, in sense of Koch & Oesterreicher (2011). This informal register is spread across different text types or genres, including expository texts (e.g., essays (20) and newspapers (21)); fiction (22); blogs about different topics (personal topics, politics, art reviews, sports, etc.), as in (23), from an educational blog, and in (24), from social media.

(20) (Silberstein 1967)

Esta fuerza de trabajo es una mercadería cualquier que
 this force of labor be.PRS.3SG a commodity.SG CUALUNQUE.SG that
 tiene un valor que está lado por el tiempo de trabajo
 have.PRS.3SG a value that be.PRS.3SG aside for the time of labor
 necesario
 necessary

‘This labor force is a commodity just like many other commodities that have a value that is set aside for the necessary labor time’

(21) Un militar de nombre cualquier, desconocido para el gran público.
 an officer of name.SG CUALUNQUE.SG unknown to the big public
 Murió el general Carlos A. Martínez.
 die.PST.3SG the general Carlos A. Martínez

‘A military officer of common name, unknown to the general public. General Carlos A. Martinez has died.’

(Télam S.E. Agencia Nacional de Noticias, 2013,
https://memoria.telam.com.ar/lesa-humanidad/202004/muri--el-general-carlos-a--mart-nez--ex-jefe-de-la-side-acusado-por-1200-cr-menes-de-lesa-humanidad_n2478)

(22) (Romano 1999: 193)

Primero apareció una ramita cualquier entre dos
 first appear.PST.3SG a twig.DIM.SG CUALUNQUE.SG between two
 ladrillos pero ahora hay toda una planta.
 bricks but now have.PRS.3SG all a plant

‘First a normal little twig appeared between two bricks, but now there is a whole plant.’

- (23) ¿qué tipo de movimiento es? Un MRUV sencillo y
 what type of movement be.PRS.3SG a MRUV simple.M.SG and
 cualquier como cualquier otro.
 CUALUNQUE.SG as any other
 ‘First a normal little twig appeared between two bricks, but now there is a
 whole plant.’
 (Blog post, http://ricuti.com.ar/no_me_salen/cinematica/c3fis_22.html)
- (24) A quién le puede interesar conocer los detalles de mi
 to whom to.him can.be.PRS.3SG interest to.know the details of my
 vida cualquier y cotidiana? La comparto como advertencia.
 life.F.SG CUALUNQUE.SG and daily.SG it be.PRS.1SG as warning
 ‘Who might be interested in knowing the details of my unremarkable and
 daily life? I share it as a warning.’
 (Twitter, 2015,
<https://twitter.com/MxAsterion/status/644188778869993474>)

The Italian form *qualunque* occurs in a more formal or written speech type in the sense of Koch & Oesterreicher (2011). Uses of *qualunque* in informal contexts are often related to the speaker’s intention to use a higher register, as in (26). In other cases, they show the speaker’s confidence with the political connotation related to Italian *uomo qualunque* ‘ordinary man’,¹⁵ as in (25).

- (25) El fascismo explícito y la reivindicación de el *uomo*
 the fascism explicit and the vindication of the.M *uomo*.SG
qualunque,
 QUALUNQUE.SG
 ‘Explicit fascism and the vindication of the “ordinary man”’
 (*Pagina/12* magazine, 2018, <https://www.pagina12.com.ar/147316-un-giro-mas-a-la-derecha-que-afecta-a-sudamerica>)

¹⁵The Italian expression *uomo qualunque* refers to an ‘ordinary or average person’, the ‘man in the street’. The term became popular in Italy in the mid-1940s, when it started being used in satirical and political magazines, in childrens’ literature, and in radio programs broadcast nationwide. However, it was first with Guglielmo Giannini’s satirical magazine *L’Uomo qualunque* and later with his political party *Fronte dell’Uomo Qualunque* that the expression rose in popularity (see Cortellazzo & Zolli 1996: 1010) and took on a political connotation referring to ‘a man that is indifferent with respect to political ideologies, especially in politics’ (see Cocco 2014: 396). Note that in our corpus the first case of *uomo qualunque* is dated 1953, a few years after Giannini’s party was founded.

- (26) Segundo, aclararle que el libro que le recomendé del
 second clarify.you that the book that to.you recommend.PFV.3SG of.the
 Cardenal Ratzinger fue escrito cuando éste era Prefecto de
 Cardinal Ratzinger was write.PTCP when this be.IPFV.3SG Prefect of
 la Congregación para la Doctrina de la Fé (no era un
 the Congregation for the Doctrine of the Faith not be.IPFV.3SG a
 teólogo qualunque por ahí), y se lo cité porque
 theologian.SG QUALUNQUE.SG for there and refl. him cite.PFV.3SG because
 Ud. afirmó y pidió alguien.
 you affirm.PFV.3SG and ask.PFV.3SG someone
 ‘Second, to clarify that the book that I recommended to you from
 Cardinal Ratzinger was written when he was Prefect of the Congregation
 for the Doctrine of the Faith (he was not just any theologian out there),
 and I quoted it to you because you affirmed and asked for someone.’
 (Blog post comment, 2011, <https://padrefabian.com.ar/la-guitarra-no-se-puede-usar-en-la-misa/#comment-3128>)
- (27) Nada bueno hay para la #Argentina en la conjunción de un
 nothing good there.is for the Argentina in the conjuncture of un
 #Qualunque y un #Maoista
 QUALUNQUE.SG and a Maoist
 ‘There is nothing good for #Argentina in the conjuncture of a
 #Qualunque and a #Maoist’ (Twitter, 2015,
<https://twitter.com/Rampa1970/status/663505695938125824>)

Plural variants have only been found in informal text forms, illustrated by examples (28–30) from Twitter.

- (28) Son re cualquieres los auriculares estos
 be.PRS.3PL very CUALUNQUE.PL the.PL headphone.PL this.PL
 ‘These headphones are very ordinary’ (Twitter, 2013,
<https://twitter.com/juanialvar3z/status/352174155883495426>)
- (29) Cuantas cualunches mal vestidas que hay en
 how.many.PL CUALUNQUE.PL bad dressed.PL that have.PRS.3SG in
 shopping [sic] por dios.
 shopping for god
 ‘How many poorly dressed ordinary people there are in the shopping
 center for God’s sake.’
 (Twitter, 2013,
<https://twitter.com/floraguerov/status/348231214219526145>)

- (30) asi es tuitar, anonimo [sic] para los cuales como
so be.PRS.3SG Twitter anonymous for the.M.PL CUALUNQUE.PL like
yo
me
‘so it’s Twitter, anonymous for common people like me’
(Twitter, 2019,
<https://twitter.com/ratakmonodosico/status/1184171952736526336>)

In texts with a greater degree of informality and colloquiality (e.g. in comment areas and social media), *cualunque* shows a greater degree of polysemy and syntactic functions, as will be described in the following sections. In these texts, it is mainly used as a Lunfardism, such as to convey irony or anger, or to build a sense of group belonging and familiarity,¹⁶ as in the following example from a local magazine blog, where the author addresses the Argentinians, including himself.

- (31) Los argentinos no estamos acostumbrados a los sabores combinados, de ahí que las especias más usadas sean la pimienta y el pimentón, a lo sumo ají molido y orégano en algunos chimichurris. Pero hay vida más allá del salero. Muchas más especias, algunas realmente exóticas que pueden hacer que un arroz cualquier sea un plato gourmet.
‘We Argentinians are not used to combined flavors, so the spices used most are pepper and paprika, at most ground chili and oregano in some chimichurris. But there is life beyond the saltshaker. Many more spices, some really exotic, that can turn ordinary rice into a gourmet dish.’
(*Perspectiva Sur*, 2016, <https://www.perspectivasur.com/3/59064-especias-esenciales-para-darle-onda-a-tus-platos>)

To summarize this section, the Arg.Sp. form *cualunque* is usually used in informal speech and not in discussions of topics of a legal or scientific nature.

It is important to note that Italian *qualunque* shows a similar register or genre restriction when it appears in postnominal position, such as *un uomo qualunque* ‘an ordinary man’, but not when it is in prenominal position, such as *qualunque*

¹⁶Teruggi (1974: 334f.) lists several reasons that can lead a person to choose a Lunfardism instead of the current Spanish word. Among these are the intention to establish contact and easy communication, as well as to express a social code. A Lunfardism can also be chosen for humorous effect or to cause irritation, to express a social criticism, to demonstrate tenderness, or even to denote scorn or contempt. For this reason, it is not uncommon to find expressive uses of Arg.Sp. *cualunque*, especially in social media.

uomo ‘any man’ (see Kellert 2021b). Kellert (2021c) argues that the correlation between register variation, position, and meaning is not surprising, given that only postnominal *qualunque* can have the evaluative meaning of ‘unremarkable’ and that expressions with this meaning can only be used in a certain register (informal) and genre type (non-scientific, non-legal genre). It seems that the most widespread form Arg.Sp. *cualunque* mimics Italian postnominal *qualunque* in this respect. This might be interpreted as a hint for the source of linguistic change (see §6).

In the following section, we look at the frequencies of the syntactic functions of *cualunque* that can give us a hint concerning its syntactic and semantic development. For this reason, we classified the collected data according to their morphosyntactic and semantic properties. As for morphosyntax, we analyzed the distribution within the nominal phrase (NP) and its co-occurrence with: a) nouns; b) determiners, such as definite and indefinite articles, possessives, and demonstratives; c) all types of quantifiers; d) prepositional phrases; e) modifiers; f) coordination with adjectives; and g) plural agreement. We also classified Arg.Sp. *cualunque* with relation to its syntactic role and its co-occurrence with: a) different types of verbs, such as transitive, intransitive, copula, and modal verbs; b) different verb moods; and c) different clause types. As for the semantic properties, we analyzed a) *cualunque*’s function and interpretation; b) the semantic features of the nouns, such as [+human] or [–human]; and c) the lexical semantics of the verbs. We then looked into correlations between different meanings and different syntactic structures. We will report the results in the following sections.

4 Frequencies of syntactic properties of Arg.Sp. *cualunque*

In this section, we will mainly deal with the frequency and the extent to which they can tell us something about the variation and change of *cualunque*. We decided to show just a few examples of each variable in the table (for more examples see Appendix A).

Table 2 shows the frequency distribution of syntactic functions of Arg.Sp. *cualunque* in our corpus. The adjective (32) is the most frequent use, compared to the noun (33) and the indefinite use as in (15), repeated below as (34). Note that the indefinite use of It. *qualunque* (see, e.g., (1), §1) is the most frequent function in Italian (Kellert 2021b).

Table 2: Frequencies of syntactic functions of Arg.Sp. *cualunque*

	n	%
Adjective	193	81%
Noun	64	18%
Indefinite	3	1%
Total	360	100%

- (32) en realidad era un lazo *cualunque*
in reality be.IPFV.3SG a bond CUALUNQUE.SG
‘it was actually a common bond’
(Blog comment, 2010,
<http://todosgronchos.blogspot.com/2010/11/todos-somos-vanesa.html>)
- (33) Ése es el *Cualunque* que designaron para dirigir la
that be.PRS.3SG the CUALUNQUE that appoint.PFV.3PL for lead the
Selección
National.Team
‘That is the ordinary person that was appointed to lead the National
Team’
(Twitter, 2017,
https://twitter.com/domingo_melia/status/945444579267969025)
- (34) Ponete una remerita *cualunque* y una calza: ¡listo!
put.IMP.2SG a shirt CUALUNQUE.SG and a leggings ready
‘Put on any shirt and a pair of leggings: ready!’
(Twitter, 2019,
<https://twitter.com/LaSuvuelabajo/status/1179938401056952320>)

We suggest the following syntactic analysis of *cualunque* in Arg.Sp. for (32) and (33).¹⁷

- (35) [Determiner *un/el* N Adj *cualunque*]
(36) [Determiner *un/el* N *cualunque*]

¹⁷As the indefinite use of *cualunque* represents only 1%, we do not analyze it in detail in this article.

Table 3 shows the frequency distribution of different positions in which the adjectival *cualunque* appears. The most frequent use is the adnominal position as in (32), followed by predicative position (37) and co-occurrence with a degree adverb (38).

Table 3: Positions of Adj *cualunque*

Positions of Adj <i>cualunque</i>	n	%
Adnominal (<i>cualunque</i> N, N <i>cualunque</i>)	261	89%
Predicative (N <i>es</i> <i>cualunque</i>)	20	7%
With degree adverbs (<i>re</i> <i>cualunque</i>)	12	4%
Total	293	100%

- (37) el ataque de Macri a Lifschitz era “cualunque”
the attack of Macri on Lifschitz be.IPFV.3SG CUALUNQUE.SG
‘Macri’s attack on Lifschitz was “ordinary”’
(*Página/12* magazine, 2016, <https://www.pagina12.com.ar/diario/suplementos/rosario/10-55616-2016-07-22.html>)
- (38) Muy *cualunque*!
very CUALUNQUE.SG
‘very ordinary!’
(Twitter, 2017,
https://twitter.com/leo_argentino/status/826620435257294850)

Table 4 shows the distribution of noun, adjective, and indefinite *cualunque* with different types of determiners (+/– definite as in, e.g., *el/un* (N) *cualunque*, quantifiers as in, e.g., *todo/algún/dos* (N) *cualunque(s)*).

The co-occurrence of Arg.Sp. *cualunque* with quantifiers and definite nouns is especially interesting, as this distribution is impossible in Italian (see Kellert 2021b). Unlike its Italian counterpart, Arg.Sp. *cualunque* can occur with definite determiners, as in (33) and (39), cardinals (40), negative quantifiers (41), existential quantifiers, such as *cualquier* ‘any’ (42) and *algún* ‘some’ (43), universal quantifiers (44), and distributive quantifiers, as in (45).

Table 4: Distribution of Arg.Sp. *cualunque* in co-occurrence with determiner types

	Adj. <i>cualunque</i>		Noun <i>cualunque</i>		Indef. <i>cualunque</i>		Total	
	n	%	n	%	n	%	n	%
Det [–def]	161	55%	20	31%	3	100%	183	51%
Det [+def]	56	19%	17	27%	0	0%	73	20%
Det [Q]	9	3%	3	5%	0	0%	12	3%
[–Det]	67	23%	24	38%	0	0%	92	26%
Total	293	100%	64	100%	3	100%	360	100%

- (39) demonstrative
 este tirano *cualunque*
 this tyrant.SG CUALUNQUE.SG
 ‘this ordinary tyrant’
 (*La Nacion*, 2018, <https://www.lanacion.com.ar/espectaculos/teatro/la-farsa-dictaduras-nid2167685>)
- (40) cardinal number
 los dos apellidos más *cualunque* de la Argentina
 the.ART.M.PL two surname.M.PL more CUALUNQUE.PL of the Argentina
 ‘the two most ordinary surnames in Argentina.’
 (Twitter, 2019,
<https://twitter.com/blackflag1974/status/1191379957919944710>)
- (41) negative quantifier
 no tolera el pelo pecho en ningún hombre
 not tolerate.PRS.3.SG the hair.M.SG chest.M.SG in any man.SG
cualunque
 CUALUNQUE
 ‘(she) does not tolerate chest hair on any ordinary man’
 (Twitter, 2012, <https://twitter.com/laspfefes/status/190825845210877953>)
- (42) existential quantifier *cualquier* ‘any’
 termino poniendo Favorito *cualquier* tweet *cualunque*.
 finish.PRS.1SG put.GER favorite any tweet CUALUNQUE
 ‘I end up bookmarking any unimportant tweet’
 (Twitter, 2012,
https://twitter.com/Palladino_shu/status/233059298383626240)

- (43) existential quantifier *algún, a* ‘some’
 Si querés saber qué anda diciendo la prensa,
 if want.PRS.2SG know.INF what go.PRS.3SG say.GER the press
 escuchá 3 minutos a algún cualquier del montón
 listen.IMP.2SG three minute.PL to some CUALUNQUE of.the pile
 ‘If you want to know what the press is saying, listen for three minutes to
 some common person of the pile’
 (Twitter, 2015,
<https://twitter.com/Bastion2008/status/655911353639636992>)
- (44) universal quantifier *todo, a*
 A todos los periodistas y/o humanos cualquieres
 to all.PL the.PL journalist.PL and/or human.being.PL CUALUNQUE.PL
 ‘To all journalists and/or common human beings’
 (Twitter, 2016, <https://twitter.com/anaopera/status/704785632200138752>)
- (45) distributive quantifier *cada* ‘each’
 Va cada qualunque, mira si no va a poder ir
 go.PRS.3SG each CUALUNQUE look.IMP.2SG if not go.PRS.3SG to be.able go
 ella?
 she
 ‘Every unimportant person goes, let’s see if she is not going to be able to
 go?’
 (Twitter, 2019,
<https://twitter.com/MarceMolt/status/1126438286564757504>)

The distribution of different quantifiers and determiners supports our syntactic analysis in (35); that is, the element *cualunque* is an adjective and not a quantifier/determiner as suggested in the literature for It. *qualunque* (see Aloni & Port 2013):

- (46) [Quantifier/Determiner *el/todo/algún* N Adj *cualunque*]

The following table summarizes the differences in the distribution of Arg.Sp. *cualunque* and It. *qualunque*. Arg.Sp. *cualunque* can be used as a (gradable) adjective. It can also be used with definite articles. These two uses are impossible in Italian (see, e.g., (5) and (7) in §1).

Let us now look at the distribution of Arg.Sp. *cualunque* in relation to verbal mood. As Table 6 shows, the indicative verbal mood, as in (32) or (45), is the most

Table 5: Uses of Arg.Sp. *cualunque* and It. *qualunque*

Language/Use	Determiner [+def]	Determiner [−def]	(Degree) Adjective
Arg.Sp. <i>cualunque</i>	+	+	+
It. <i>qualunque</i>	−	+	−

frequent one (see Appendix B for further examples).¹⁸ Uses under the subjunctive, as in (47), the imperative, as in (34), or modal verbs, as in (48),¹⁹ are much less frequent.

(47) (Medina 1989: 287)

Si vos fueras un *cualunque*
if you.SG be.SBJV.2SG a CUALUNQUE
'If you were a common person'

(48) no se puede ver ni un noticiero *cualunque*
not REFL can.PRS.3SG see even a newscast CUALUNQUE
'you can't even watch a regular newscast'
(Twitter, 2018, <https://twitter.com/Buda48/status/1055953281929920522>)

Table 7 shows the distribution of Arg.Sp. *cualunque* with respect to verbal mood indicative and tense/aspect (see Appendix B for further examples). The most frequent use of *cualunque* is with the indicative present tense, as in (49), and the past tense with perfect and imperfect aspect, as in (50) and (51), respectively.

(49) No tenés una parrilla *cualunque*?
no have.PRS.2SG a grill CUALUNQUE
'Don't you have a common grill?'
(Twitter, 2018,
<https://twitter.com/blackflag1974/status/1022158271384379392>)

(50) (Marchetti 2014)
Me compré uno de esos jugos *cualunque*
REFL buy.PFV.1SG one of those juice.PL CUALUNQUE.PL
'I bought one of those bad juices'

¹⁸Sentences with verbal ellipsis have not been counted. The number of cases with verbal ellipsis amounts to 95, constituting 26% of the total number of occurrences.

¹⁹Cases with structures of the type [Vmod Vinf] and [Vfin Vinf] have also been counted under [infinitive]. They constitute 23% and 15% of [infinitive] cases, respectively, with 9 [Vmod Vinf] and 6 [Vfin Vinf] occurrences out of 40.

Table 6: Arg.Sp. *cualunque* and verbal mood

V mood	n	%
Indicative	196	74%
Infinitive	40	15%
Subjunctive	13	5%
Gerund	9	3%
Imperative	4	2%
Conditional	3	1%
Total	265	100%

- (51) (Lozza & Pugliese 1985: 45)
Iba con un traje *cualunque*.
go.IPFV.3SG with a suit CUALUNQUE
‘He was wearing a cheap suit.’

Note that the distribution of verbal mood, tense, and aspect that co-occurs with *cualunque* stands in contrast to the distribution of verbal mood, tense, and aspect of Free Choice elements such as the prenominal It. *qualunque* or Engl. *any*, which are only possible in modal contexts and are ungrammatical in non-modal contexts such as with verbs with perfective aspect, indicative mood present, and past tense (see Quer 2000, Aloni & Port 2013, Giannakidou & Quer 2013, Kellert 2021a). We thus suggest that Arg.Sp. *cualunque* has undergone a process of word class change that is visible not only with respect to the type of Determiner Phrase (DP), but also with respect to Tense and Aspect. This change needs to be explained (see §5 and §6).

To sum up, we have shown major differences in the use of Arg.Sp. *cualunque* and It. *qualunque* which are: a) Arg.Sp. co-occurs with definite determiners, and quantifiers, b) it can be used as a gradable adjective, and c) it is not restricted to modal contexts, but often appears with indicative mood in the present or past tenses (with perfective aspect).

5 Semantic analysis of *cualunque*

In this section, we suggest a detailed semantic analysis of Arg.Sp. *cualunque*. We first start with the notion of the anti-specificity of indefinites and quantifiers in order to explain how this notion is related to It. *qualunque* or similar items in Italian dialects that were used by Italians who emigrated to Argentina. We will

Table 7: Distribution of Arg.Sp. *cualunque* with relation to indicative V tense

Indicative V tense and aspect	n	%
Present	142	72%
Perfect	31	16%
Imperfect	20	10%
Future	3	2%
Total	196	100%

then postulate that a sort of pragmatic strengthening happened in later periods that has induced a semantic and syntactic change of *cualunque* in Arg.Sp., which has been described in the present section.

5.1 Anti-specificity

When governed by a modal sentence, Free Choice indefinites such as It. *qualunque* are interpreted as triggering a set of alternatives (see §1). That is, under modals, *qualunque libro* or *un libro qualunque* denotes a set of possible books, such as this book or that book (see §1). The phrase *qualunque libro* or *un libro qualunque* ‘any book’ does not refer to a specific book, in contrast to *questo libro* ‘this book’ or *il libro che mi hai portato ieri* ‘the book you brought to me yesterday’. This contrast of +/- specific nouns has been already observed in the literature (see von Heusinger 2011, Etxeberria & Giannakidou 2014, Giannakidou & Quer 2013 and references therein). Specificity is indicated when “a speaker uses an indefinite noun phrase and intends to refer to a particular referent” (von Heusinger 2011: 10). Conversely, referential vagueness can be interpreted as an absence of specificity (von Heusinger 2011, Giannakidou & Quer 2013). Free Choice indefinites, such as *qualunque* and *cualquiera*, and epistemic indefinites, such as *algún* in (53), are anti-specific and express referential vagueness. They thus signal a lack of referential intent (von Heusinger 2011). Referential vagueness can be described as a variation in possible values of the expression and in uncertainty about which one is the actual value. The uncertainty about the actual value follows from the Free Choice interpretation, which states that any value is possible (see Aloni & Port 2013, among many others).

One way to test referential vagueness and anti-specificity is to use the specificity test, in which the referent of the indefinite noun is specified post-hoc, as in the *it*’s-clause in (52):

- (52) There is someone at the door. It's Mr. Smith.

Such continuations are incompatible with *algún*-phrases, as illustrated in (53) (see Menéndez-Benito 2010, Alonso-Ovalle & Menéndez-Benito 2010):

- (53) (Alonso-Ovalle & Menéndez-Benito 2013: 36)
 María se casó con algún médico. # En concreto con el Dr. Smith.
 'María married some doctor. Concretely, Dr. Smith.'

Free Choice indefinite phrases are also incompatible with continuations, such as (53) (see Chierchia 2006, Jayez & Tovenia 2007, Fălăuș 2014).

Anti-specific indefinites never modify singleton nouns, that is, nouns that denote a single atomic individuum such as a single person or thing (the so-called "anti-singleton condition", cf. Menéndez-Benito 2010). They must occur with nouns that denote a set with at least two members. This prediction follows from the Free Choice interpretation of *qualunque*, which entails a consideration of at least two alternatives. This prediction is borne out empirically in our data, where N *qualunque* does not occur with singleton nouns, as shown in (54). Here, N *qualunque* entails at least two different individuals with the name Angela Merkel and at least two second world wars. However, this use is incompatible with our world, in which Angela Merkel refers to a single person, namely the German chancellor and in which the Second World War refers to a single war that occurred from 1940 to 1945:

- (54) # *un, a* singleton N *qualunque* 'any singleton N' (# means 'awkward')
 a. # *una Angela Merkel qualunque* 'any Angela Merkel'
 b. # *una Segunda Guerra Mondiale qualunque* 'any second world war'

Let us now see whether Arg.Sp. *cualunque* can be described with the notion of anti-specificity. Recall that in modal contexts such as imperatives, *cualunque* can have the Free Choice interpretation of 'any'; that is, every alternative is possible, and the speaker does not distinguish between some alternative or another. *Cualunque* as Free Choice is anti-specific:

- (55) Ponete una remerita *cualunque* y una calza: ¡listo!
 put.IMP.2SG a shirt CUALUNQUE and a leggings ready
 'Put on some shirt and a pair of leggings: ready!'
 (Twitter, 2019,
<https://twitter.com/LaSuvuelabajo/status/1179938401056952320>)

- (56) Almorzamos en una pizzería cualquier y comemos postre en la eat.PRS.1PL in a pizzeria CUALUNQUE and eat.PRS.1PL dessert in the casa de los azulejos, porque somos peronistas. house of the azulejos because we are peronistas.
 ‘Let’s eat in a pizzeria no matter which one and eat dessert in the house of azulejos, because we’re Peronistas (= member of a particular political party).’
 (Twitter, 2016, https://twitter.com/di__tir/status/813118305548570624)

In (21), repeated below as (57), *cualunque* is interpreted as ‘epistemically unknown’ (see Chierchia 2013 on a similar kind of interpretation with It. *qualunque*):

- (57) Un militar de nombre cualquier, desconocido para el gran público. Murió el general Carlos A. Martínez.
- a. ‘A military officer of some name, unknown to the general public. The general Carlos A. Martínez has died.’
 - b. ‘There is a military officer with a name that is unknown to the general public. This name is Carlos A. Martínez. He has died.’
- (Télam S.E. Agencia Nacional de Noticias, 2013, https://memoria.telam.com.ar/lesa-humanidad/202004/muri--el-general-carlos-a--mart-nez--ex-jefe-de-la-side-acusado-por-1200-cr-menes-de-lesa-humanidad_n2478)

The sentence in (57b) states that according to the general public (i.e. the agent of the epistemic modal base),²⁰ the officer’s name is unknown. From this it follows that the name can be every possible name. Thus, all names are the same, according to the general public’s knowledge. The Free Choice interpretation of *cualunque* is compatible with the ignorance interpretation indicated by *desconocido* ‘unknown’ given in the context of this example.

To sum up, Free Choice indefinites and epistemic indefinites are anti-specific, referentially vague, and never occur with singleton nouns.

As we will see in §5.2, Free Choice Indefinites can lead to a *pragmatic strengthening* (i.e. strengthening the unspecific meaning of *cualunque*), which can eventually lexicalize and lead to semantic change. In §6, we will claim that this is exactly what happened with Arg.Sp. *cualunque*.

²⁰The difference between this interpretation and the “epistemic unknown” interpretation of It. *qualunque* suggested in Chierchia (2013) is that in Chierchia’s analysis, the epistemic unknown refers to the speaker and not to the agent as is the case with Arg.Sp. *cualunque*.

5.2 Pragmatic strengthening

We saw in §4 that *cualunque* is used as an adjective and can have the neutral interpretation of ‘ordinary’, ‘common’, or ‘normal’, as in (58) or (59):

- (58) si sos cliente Black te atienden inmediatamente, si
if be.PRS.2SG client Black you attend.PRS.3SG immediately, if
sos *cualunque*, esperà [sic] dos horas.
be.PRS.2SG CUALUNQUE, wait.PRS.3SG two hours.
‘if you are the client Black they attend to you immediately, if you’re an
ordinary person, you have to wait for two hours.’
(Twitter, 2013, <https://twitter.com/SuHerBre/status/292399334497058816>)
- (59) un dibujo clásico, tradicional, *qualunque*.
‘a classic, traditional, ordinary drawing’
(Blog comment, 2013, <http://www.comiqueando.com.ar/secciones/el-podcast-de-comiqueando/programa-52/>)

The example of *cualunque* in (58) is used as a predicate describing clients that do not have any particular property and thus represent a kind of person that stands in opposition to a certain and particular type of client, namely *cliente Black*. The example *un dibujo cualquier* in (59) refers to a kind of drawing described as classic or traditional that does not have any particular properties distinguishing it from very unusual or modern drawings (e.g., drawings by Picasso). To summarize what has been said so far, N *cualunque* refers to a kind of people or object that does not have any specific or particular properties that would distinguish it from others of the same kind. This kind stands in a contrastive relation with another kind that describes people or objects that have particular or distinguishing properties such as *cliente Black* or drawings by Picasso.

The neutral meaning of N *cualunque* is used in a situation where *cualunque* refers to a set of individuals that do not have any distinguished properties. These individuals with no distinguished properties are evaluated somewhere in the middle of a scale that ranks individuals according to their distinguishing properties. For instance, *cualunque* in (59) refers to classic drawings that are opposed to distinguished drawings, which can be either especially good or especially bad (see also Alonso-Ovalle & Royer (2021) for an implicit ranking scale of *komon* on the ‘unremarkable’ reading in the Mayan language Chuj).

The neutral meaning of N *cualunque* often occurs when *cualunque* is evaluated on a frequency scale. The *cualunque*’s meaning as ‘common’ in *un día cualquier* refers to some property of the day, which is very typical for many days and is

thus very frequent, such as those filled with some usual activities such as waking up in the morning, going to work, and going to bed in the evening. This kind of a day is contrasted to some specific or particular day, such as a birthday, which may include unusual activities such as celebrating and staying home from work. Particular or specific days are rare because they have certain features that distinguish them from common days. Actually, all examples in the corpus with *N cualquier* that contain a noun denoting some expression of time, such as *domingo* or *día* as in (60), can be interpreted as situated in the middle of a frequency scale.

- (60) Noooo Bipolarlo es lo mejor de este domingo cualquier.
 Noooo Bipolarlo be.PRS.3SG the best of this sunday CUALUNQUE.
 Tiene la pluma de oro el que redacta esos
 hold.PRS.3SG the pen of gold the that edits these
 ‘No Bipolarlo is the best of this ordinary Sunday. The one who edits those tweets has the Golden Pen.’
 (Twitter, 2018,
<https://twitter.com/caradecumbiaok/status/1046464719530467328>)

Another example with the neutral meaning of *cualunque* as ‘common’ is given in (61), where the relative clause *camina por la calle* restricts the set of people to those who walk on the street and who can be observed very frequently. This relative clause makes thus reference to normal or common people:

- (61) Al porteño cualquier que camina por la calle
 the citizen of Buenos Aires CUALUNQUE who walk.PRS.3SG on the street
 no le interesa lo de la base militar china.
 not him interest.prs.3sg the of the base military Chinese
 ‘The common citizen of Buenos Aires who walks on the street is not interested in the Chinese military base.’
 (Twitter, 2015, <https://twitter.com/ton011972/status/562337257328508929>)

The common interpretation with reference to people is often given when the speaker refers to many people, including himself as *como nosotros* ‘like us’ in (62) or as *como yo* ‘like me’ in (63). Note that in the latter example, *cualunque* is used in contrast to a distinguished type of individuals: *gente de clase alta*.

- (62) [...], bardear a personas *cualunque* como nosotros en una red
to.fence to people CUALUNQUE like us in a network
social por opinar distinto es muy pedorro.
social for to.think different be.PRS.3SG very annoying
‘[...], to fence normal people like us in a social network for having a
different opinion is very annoying.’
(Twitter, 2019,
https://twitter.com/Felicitas73_/status/1174698124197212161)
- (63) Y no hablo de políticos o gente de clase alta, hablo
I not talk.PRS.1SG about politicians or people of class high, talk.PRS.1SG
de *cualunque* como yo
about CUALUNQUE like me
‘I’m not talking about politicians or upper-class people, I’m talking about
common people like me’
(Twitter, 2018, <https://twitter.com/KaroSci/status/1011244306898866176>)

The same analysis can be applied to *un dibujo cualquier* in (59). Classic or traditional drawings are more common in this speech context than drawings with certain properties that are distinguished from classic or traditional drawings (e.g., drawings by Picasso). Clients with undistinguished properties referred to by *cualunque* in (58) are more common than those with distinguished properties such as *cliente Black*.

To sum up, the ‘common’ meaning of *cualunque* is the result of considering individuals denoted by N *cualunque* as being in the middle of some scale, such as a scale of frequency or some other scale and the result of contrasting these individuals to particular and rarely observable individuals.

The derogatory meaning is the result of evaluating the qualities of the kind of individuals denoted by *cualunque* in contrast to specific individuals with positive qualities. Thus, the evaluation of *cualunque* shifts from the middle of a scale to an extreme end as the result of contrasting *cualunque* to a set of especially good individuals. The derogatory meaning of *cualunque* in (64) and (65) introduces a contrast between exceptional people such as *el actor original* or *periodista operadora* and unexceptional people with no distinguished qualities or a simple activist:

- (64) Pero que vuelva el actor original, no este cualquier
 but that return.SBJV.3SG the actor original, not this CUALUNQUE
 sacado de Antares.
 take.PST.PTCP from Antares
 ‘But let the original actor return, not this poor one taken from Antares.’
 (Twitter, 2019, <https://twitter.com/AlanG996/status/1115053249197953024>)
- (65) Manguel cayó de periodista operadora K a simple militante
 Manguel fall.PST.3SG from journalist operator K to simple militant
 cualquier. La degradaron.
 CUALUNQUE. Her degrade.PST.3PL
 ‘Manguel fell from journalist operator K to simple ordinary activist. She
 was demoted.’
 (Twitter, 2019,
https://twitter.com/La_gringai/status/1181608328549584896)

The same shift towards a negative meaning of ‘common/normal/usual’ can be observed in diachrony. We find lexicalized pejorizations of the meaning ‘common’ or ‘average’ in other languages such as Fr. *vulgaire*, Engl. *vulgar* or Fr. *médiocre* or Sp. *mediocre* (see Kleparski & Borkowska 2007: 43).

We leave it open for future research why *cualunque* is never contrasted to specific individuals or types that are exceptionally bad on the non-neutral interpretation. One could possibly derive this fact from the Gricean maxim of *informativity* (see Geurts 2010, Rosemeyer, p.c.). Thus, *cualunque* is not contrasted with bad because such a contrast would not be informative enough according to Gricean maxim of informativity. This is why we do not hear sentences like “he is not a bad writer, but an ordinary writer,” because being an ordinary writer does not contrast sufficiently enough being a bad writer. However, the contrast between a good writer and an ordinary writer is informative enough, because being ordinary can mean being bad in certain contexts. This is why we do hear sentences like “he is not a good writer; he’s an ordinary writer.”²¹

However, it seems to be a very general pattern that properties describing singleton nouns such as *unique*, *particular*, *outstanding* and *special* have a positive meaning more often than a negative one; for instance, a *grade with distinction* is an especially good grade. *John is a remarkable man* means that John has some positive properties that distinguish him from other men. *An extraordinary day* is generally an extremely good day and not an extremely bad day unless uttered

²¹We thank Malte Rosemeyer for discussing this point with us.

with a special intonation. This might explain why *cualunque* is usually contrasted with exceptionally good individuals on the non-neutral meaning.

To sum up so far, Arg.Sp. *cualunque* and It. *qualunque* can have the neutral meaning of ‘common’ as ‘widespread’ or some pejorative version of ‘common’ with the meaning ‘common’ as ‘worse than extraordinary’. We analyzed these meanings as a property over (kinds of) individuals, which imply different scales: a frequency scale and a scale of goodness.

6 Hypothesis on meaning change

We assume that *cualunque* underwent a change into a degree predicate *cualunque* in (66) and a noun in (67) from postnominal indefinite *cualunque*:

- (66) It. postnominal indefinite *qualunque* (original construction) >> Arg. Sp. postnominal *cualunque* (by lexical borrowing from Italian) >> degree adjective *cualunque* (by pragmatic strengthening and syntactic recategorization)
e.g., *cualunque* ‘any’ >> (re/tan) *cualunque* ‘very ordinary/bad’
- (67) *un,a qualunque* >> *un,a cualunque* (lexical borrowing from Italian) >> *el/la cualunque* (semantic shift/recategorization)
- a. las leyes son para los *cualunque*, o sea nosotros.
the laws be.PRS.3SG for the CUALUNQUE, or be.SBJV.3SG us
‘the laws are for normal people, that is, for us.’
(Twitter, 2019,
<https://twitter.com/maggiopalacios4/status/1180456730028851201>)
- b. Este *cualunque* que se cree presidente hasta cuando
this CUALUNQUE who REFL believe.PRS.3SG president until when
hay que soportar sus mentiras es un payaso como
have.PRS.3SG that support his lies be.PRS.3SG a clown like
Maduro de Venezuela
Maduro of Venezuela
‘This ordinary person that believes himself to be president until you have to believe his lies is a clown like Maduro of Venezuela’
(Twitter, 2019,
<https://twitter.com/Alfredo00649870/status/1085118326324514816>)

In order to explain the shift of *cualunque* into a degree adjective as represented in (66) or as an evaluative noun in (67), we assume that the scale of goodness,

which implies a degree scale (i.e., good is gradable), has been lexicalized in Argentinian Spanish. The lexicalization of the degree scale has driven the reanalysis of *cualunque* as a degree adjective or as a nominal element with an evaluative interpretation:

- (68) [DegP *re/muy* ‘very’ [Adj *cualunque*]] ‘very ordinary/bad’
(69) [DP [Noun [+eval] *cualunque*]] ‘someone ordinary/bad’

As already mentioned in §1, diachronic sources are very scarce. The first diachronic occurrences of *cualunque* in texts written in literary Cocoliche (see 70–72) have a Free Choice interpretation:

- (70) Cocoliche
Che cosa volete? – [...]. – Cualunque cosa,-
what want.PRS.2PL CUALUNQUE thing
‘What do you want? – [...] – Any thing’
(*La Mujer* magazine, 1900,
<https://archive.org/details/lamujer2140unse/page/n191?q=cualunque>)
- (71) Cocoliche
Lu arquila per dos peso in cualquier montepío.
it.ACC pawn.PRS.3SG for two pesos in CUALUNQUE pawnshop
‘He pawns it for two pesos in any pawnshop.’
(*P.B.T.* magazine, 1906,
<https://digital.iai.spk-berlin.de/viewer/fullscreen/861383842/105/>)
- (72) Cocoliche
Il cochiyo, la fareñera, la finyinga, [...] cualquier de cueli [...] te
the knife the dagger the stab CUALUNQUE of these you
vale mase que il cuore
be.worth.3SG.PRES more than the heart
‘The knife, the dagger, the stab, any of these are worth more than the heart’
(*Atlántida* magazine, 1930,
<https://books.google.com.ar/books?id=cB8QAAAAIAAJ>)

Finally, the degree predicate *cualunque* and the string [definite/demonstrative N *cualunque*] represent a recent development in Argentinian Spanish. Younger people are more likely to use *cualunque* as a degree adjective than older people, as our investigation of tweets and a linguistic poll on the Facebook group members

of Lingüística Argentina has shown (Kellert 2021a). This sociolinguistic contrast can be seen as one indicator of gradual language change (see Stein 1990, Seiler et al. 2018).

We assume that the syntactic change schematized in (67) is a consequence of the lexicalization of the pragmatic scale on which speakers and/or hearers evaluate individuals denoted by *N cualquier* in contrast to individuals with specific or special properties. This kind of process of lexicalization of pragmatic inferences is known in the literature on semantic change as *pragmatic strengthening*, described as the process by which meanings tend to be enriched in pragmatic contexts and the resulted implicatures can be eventually conventionalized. This conventionalization is what triggers semantic change (see Traugott 1989: 35, among others). Moreover, Traugott observes that “meanings tend to become increasingly based in the speaker’s subjective belief state/attitude toward the proposition” (Traugott 1989: 35). This process is known in the literature as *subjectification* or *pragmaticalization* (see (Diewald 2011)). We believe that *cualunque* is another example of pragmatic strengthening, subjectification, and pragmaticalization, because its new meanings are the results of contextual interpretation by the speaker and hearer and the evaluation of individuals on different scales.²²

Our analysis of *cualunque* denoting a property with different readings such as ‘common’ and the derogatory version of it as ‘not outstanding’ can account for the adjectival behavior of *cualunque* observed in §4.

7 Summary and outlook

We have analyzed the synchronic variation of *cualunque* in Arg.Sp. and *qualunque* in Italian and have identified different syntactic categories of *cualunque* with different interpretations:

²²The exact syntactic analysis of the postnominal indefinite *cualunque* in (66) (i.e. the original construction) needs to be studied in the future (see Kellert 2021a). One possibility is to assume a two-determiner-analysis of UN *N cualquier*, as has been suggested for UN *N qualunque* (see Zamparelli 2000, among others). On these accounts, *qualunque* is considered to be a strong quantificational determiner on the same lines as ‘every’, ‘some’, or ‘none’. As a consequence, the indefinite determiner *un* in *un N qualunque* is analyzed as an empty or weak element with no semantic value. One weak point of the determiner analysis of *qualunque* is that the change of Free Choice indefinites into degree predicates or evaluative nouns would be a case of degrammaticalization (see Kellert 2021a), because determiners are functional/grammatical categories that usually do not change into lexical categories. Usually, it is the opposite that occurs; that is, lexical categories change into functional/grammatical categories (see Roberts & Roussou 2003). Another weak point of the determiner analysis is that determiners in Romance languages are usually used in prenominal positions, rarely in postnominal positions (see Stark 2006 for some exceptions).

Cualunque as a Free Choice indefinite ‘any’ or as an epistemic indefinite ‘some’ with the meaning ‘epistemically unknown’.

Cualunque as a (gradable) adjective with the neutral meaning of ‘common’ as ‘widespread’ or ‘frequent’ or ‘normal’ or with the derogatory meaning of ‘common’ as ‘worse than outstanding’ or even ‘bad’. This syntactic use is not possible with Italian *qualunque*, although both interpretations are possible in Italian in the predicative context of *UN N qualunque*.

Cualunque as a noun with either a neutral or depreciative meaning. The nominal use is only possible with indefinite determiners in Italian, not with definite determiners or quantifiers as in Argentinian Spanish.

The analysis suggested in this article can be applied to similar cases in other Romance languages that allow similar indefinites in non-modal contexts such as predicative position. This is shown for French *n’importe quoi*, *quelconque*, and *cualquiera* in Argentinian Spanish or the nominalized *cualquiera* in European Spanish (see Kellert 2021a, see also Kellert & Enrique-Arias 2024 [this volume] for Catalan *qualsevol*):

- (73) French
C’est du n’importe quoi.
it.be.PRS.3SG of.the N’IMPORTE QUOI
‘It’s total nonsense.’
- (74) French
C’est très quelconque.
it.be.PRS.3SG very QUELCONQUE
‘It’s very ordinary.’
- (75) Argentinian Spanish
Es (re) cualquiera.
be.PRS.3SG RE CUALQUIERA
‘It’s really worthless/nothing special’
- (76) European Spanish
Juan es un cualquiera.
Juan be.PRS.3SG a CUALQUIERA
‘Juan is a nobody/low-class person.’

In the future, we will study the question what determines the pejorative or the neutral interpretation of *cualunque*. So far, it seems that one important feature

that biases one interpretation or the other is the inclusion or exclusion of the speaker in the set of ordinary people denoted by the meaning of *N cualquier* (see §6). If the speaker includes herself in the set of ordinary people, *cualunque* has a neutral meaning, whereas if the speaker does not, then *cualunque* has only the pejorative interpretation:

- (77) (exclusion of the speaker) pejorative meaning
 sos una cualquier.
 be.PRS.2SG a CUALUNQUE
 ‘You are an unimportant/low-class woman’
 (Twitter, 2013,
<https://twitter.com/BrendaCapello/status/369843684021252096>)
- (78) (inclusion of the speaker) neutral meaning
 las leyes son para los cualquier, o sea nosotros
 the laws be.PRS.2PL for the.2PL CUALUNQUE, or be.SBJV.3SG us
 ‘the laws are for normal people, that is, for us’
 (Twitter, 2019,
<https://twitter.com/maggiopalacios4/status/1180456730028851201>)

Another important factor that biases the pejorative use is the lexical semantics of the noun modified by *cualunque*. If the noun is a depreciative word, like *negro* in (79) (see Kellert 2021d), and the speaker excludes herself from the set, *cualunque* has a pejorative interpretation:

- (79) Sos un negro cualquier.
 be.PRS.2SG a NEGRO CUALUNQUE
 ‘You’re a (just) a simple person/You’re nobody important’ (pejorative use)
 (Twitter, 2020,
<https://twitter.com/LeluuArtero/status/233997516054601728>)

If the same lexical noun co-occurs with positive expressions such as *te quiero* ‘I love you’, this creates an ironic meaning due to the use of depreciative nouns like *negro cualquier* in a positive context:

- (80) luchiditatta si8siis te quiero negrito cualquier
 luchiditatta si8siis you love.PRS.1SG NEGRITO CUALUNQUE
 ‘I love you little unimportant person’
 (Twitter, 2014, <https://twitter.com/facuundit/status/438433004042846208>)

- (81) Nacho: Te amo negra cualunque♥
 Nacho: you love.PRS.1SG NEGRA CUALUNQUE
 ‘Nacho: I love you unimportant female person’
 (Twitter, 40, <https://twitter.com/MelaCerioli/status/431786831042994176>)

Another observation is that scalar focus particles can bias the derogatory meaning of *cualunque* as well (see Kellert 2021a). The focus particle *solo* ‘only/mere’ has the function of excluding high-value alternatives. In the following example, the focus particle *solo* associates with a focus alternative *una banda cualunque*, which leads to the exclusion of all other alternatives (‘particular bands’). This exclusion leads to the interpretation of *banda cualunque* as a band being low on the scale of high-quality bands:

- (82) Pense que el sabado ibamos a ir a una re fiesta, y al
 think.PST.1SG that the saturday go.PFV.1PL to go to a very party, and in
 final solo toca una banda cualunque.
 end only play.PRS.3SG a band CUALUNQUE
 ‘I thought that on Saturday we were going to a good festival, but in the
 end only an ordinary band was playing.’
 (Twitter, 2016,
<https://twitter.com/solariascelli/status/722543527159324672>)

An interesting shift in semantic interpretation can be observed with pets (see 83). There, the meaning of *cualunque* is ‘mixed-breed’, e.g., *gato cualunque* (84) ‘mixed-breed cat’ or *perro cualunque* (85) ‘mongrel’. We assume that this semantic shift is probably due to the meaning of *cualunque* as ‘not special/not outstanding’. In the domain of pets, the meaning ‘not-outstanding’ means not belonging to a ‘pure breed’.

- (83) Mis bebas son cualungues/mestizas osea no son
 my babies be.PRS.3SG CUALUNQUE/half-breed/mixed, i.e., not be.PRS.3PL
 de raza solamente mi ahijado [...] que se llama León
 of pedigreed, only my godson [...] that REFL be.called.PRS.3SG Leon
 y es un chihuahua.
 and be.PRS.3SG a chihuahua.
 ‘My babies are half-breed/mixed, i.e., they’re not pedigreed, only my
 godson [...] whose name is Leon and who’s a chihuahua.’
 (Twitter, 2018,
<https://twitter.com/maruuchis85/status/968997312113586176>)

- (84) (Medina 1989: 132)
 luego volvió a te con un gato cualquier
 later return.PST.3SG to you with a cat CUALUNQUE
 ‘(he) later came back to you with a mixed-breed cat’
- (85) A mi me gustan los perros cualquieres.
 to me me like.PRS.3PL the dogs CUALUNQUE.PL
 ‘I like mongrels.’
 (Twitter, 2017, <https://twitter.com/QmaxiQ/status/892576809400848387>)

This hypothesis needs to be checked in future research. In the future, we will provide a detailed semantic analysis of the evaluative meaning of *cualunque* (see Gutzmann 2013’s analysis of expressive elements).

Acknowledgements

We thankfully acknowledge the funding provided by the Deutsche Forschungsgemeinschaft (DFG) for the project “Quantification in Old Italian”. A warm thanks goes to Andrés Saab and all our informants for their valuable comments on Argentinian Spanish.

Appendix A Data reflecting Table 3 in §4

- (86) [cualunque N] (Boot 2012)
 un cualquier maestro mayor de obra.
 a CUALUNQUE master senior of work
 ‘a general contractor like any other’
- (87) [N cualquier]
 un programa cualquier de televisión
 a program CUALUNQUE of television
 ‘an ordinary TV program’
 (*Rock and Ball* magazine, 2013, <https://rockandball.com.ar/punto-de-vista/futbol-para-todos-menos-para-los-que-van-a-la-cancha-76096/>)
- (88) [N parece cualquier]
 me pareció re cualquier la interpretación
 to.me appear.PFV.3SG very CUALUNQUE the interpretation
 ‘I found the interpretation really bad’
 (Twitter, 2015, <https://twitter.com/LukeAKD/status/605901400157978624>)

- (89) [N es cualquier]
tus transmisiones de fútbol son ordinarias y cualquieres
your broadcast.PL of soccer be.PRS.3PL ordinary.PL and CUALUNQUE.PL
‘your soccer broadcasts are ordinary and of low quality’
(Twitter, 2016, <https://twitter.com/CCesaroni/status/696465325131460608>)
- (90) [DegP muy ‘very’ [Adj cualquier]]
Día muy cualquier en la oficina.
day very CUALUNQUE in the office
‘a very ordinary day at the office’
(Twitter, 2012,
<https://twitter.com/IvanDawidowski/status/217632663304011776>)
- (91) [DegP re ‘very’ [Adj cualquier]]
un tag re cualquier
a tag very CUALUNQUE
‘a very ordinary tag’
(Twitter, 2016,
<https://twitter.com/S4NFR4NC15C0/status/748491445858934785>)
- (92) [DegP tan ‘so’ [Adj cualquier]]
El día que nació fue tan cualquier que [...]
the day that be.born.PFV.1SG be.PFV.3SG so CUALUNQUE that
‘the day I was born was so ordinary that [...]’
(Twitter, 2013,
<https://twitter.com/casicasiperono/status/371041190012919808>)
- (93) [DegP más o menos ‘more or less’ [Adj cualquier]]
en un restaurante más o menos cualquier
in a restaurant more or less CUALUNQUE
‘in a more or less ordinary restaurant’
(Twitter, 2018,
<https://twitter.com/tamtenenbaum/status/1048597979416944640>)
- (94) [DegP medio ‘half’ [Adj cualquier]]
pero ese me parece medio cualquier... o no?
but that to.me seem.PRS.3SG half CUALUNQUE or not
‘but that one seems kind of ordinary... or does it?’
(Twitter, 2014,
<https://twitter.com/GuilleSandrini/status/442802662879137792>)

Appendix B Data reflecting Table 6 and Table 7 in §4

- (95) indicative present
 Pero, yo no soy un esquiador *cualunque*.
 but I not be.PRS.1SG a skier CUALUNQUE
 ‘but I am no ordinary skier’
 (*La púrpura de tiro*, 2019, <https://www.lapurpuradetiro.com.ar/index.php/numeros-anteriores/item/1111-lo-siento-senor-griggs>)
- (96) indicative past perfective
 la crisis [...] se trasladó al pueblo, a los *cualunque*s
 the crisis REFL pass.PFV.3SG to.the people to the CUALUNQUE.PL
 ‘the crisis [...] passed to the people, to the lower-class people’
 (*Tiempo Argentino*, 2018, <https://www.tiempoar.com.ar/politica/damian-selci-el-militante-es-el-producto-mas-civilizado-que-puede-tener-una-sociedad/>)
- (97) indicative past imperfective
 mirá como estaban ahí los *escribas* con una
 look.IMP how be.IPFV.3PL there the.PL scribe.PL with a
llamita *cualunque*
 flame.DIM.F.SG. CUALUNQUE
 ‘look how the scribes were there with a common little flame’
 (Twitter, 2016, https://twitter.com/vanesagiselle_/status/786411452735352832)
- (98) indicative future (Asis 2000: 82)
 el resto de su vida será un vendedor *cualunque*.
 the rest of his life be.FUT.3SG a salesman CUALUNQUE
 ‘the rest of his life he will be an ordinary salesman.’
- (99) modal verb *deber* ‘must’
 ¿Se debe comparar a un equipo *cualunque* con una
 REFL should.PRS.3SG compare to a team CUALUNQUE with a
maquinaria ideal y perfecta como es el Barcelona?
 machine ideal and perfect as be.PRS.3SG the Barcelona
 ‘Should we compare an ordinary team with an ideal and perfect machine like Barcelona?’
 (*Clarín* magazine, 2016, https://www.clarin.com/opinion/mania-discutirlo_0_EkX9qo-XW.html)

- (100) infinitive
 Me enferma oir [sic] a cualesques tratar de parecer
 me make.sick.PRS.3SG hear.INF to CUALUNQUE.PL try.INF of seem
 finas hablando con arrastre.
 elegant.PL speaking with affectation
 ‘It makes me sick to hear low-class people trying to look elegant
 speaking in an affected way.’
 (Blog post, 2018,
<http://nanopoder.blogspot.com/2008/04/cosas-odiosas.html>)
- (101) gerund
 Te estas [sic] comiendo una [sic] postre cualquiere de chocolate
 you be.PRS.2SG eat.GER a dessert CUALUNQUE of chocolate
 simil serenito. No te hagas....
 similar cool not you do.SBJV.2SG
 ‘You’re eating a shitty chocolate dessert like nothing. Don’t do it...’
 (Twitter, 2018,
<https://twitter.com/prestoyvoila/status/960917953674956800>)
- (102) conditional
 con Suarseneguer u otro sería una peli cualquiere
 with Schwarzenegger or other be.COND.3SG a film CUALUNQUE
 ‘with Schwarzenegger or someone else would be an ordinary movie’
 (Twitter, 2019,
<https://twitter.com/DiegolBarraza/status/1177618733788995584>)

References

- Aloni, Maria & Angelika Port. 2013. Epistemic indefinites crosslinguistically. In Yelena Fainleib, Nicholas LaCara & Park Yangsook (eds.), *Epistemic indefinites crosslinguistically*, 29–43. Amherst, MA: GLSA Publications.
- Alonso-Ovalle, Luis & Paula Menéndez-Benito. 2010. Modal indefinites. *Natural Language Semantics* 18(1). 1–31. DOI: 10.1007/s11050-009-9048-4.
- Alonso-Ovalle, Luis & Paula Menéndez-Benito. 2013. A note on the derivation of the epistemic effect of Spanish Algún as an implicature. In Anamaria Fălăuş (ed.), *Alternatives in semantics*, 36–49. London: Palgrave Macmillan UK. DOI: 10.1057/9781137317247_2.

- Alonso-Ovalle, Luis & Paula Menéndez-Benito. 2017. Epistemic indefinites: On the content and distribution of the epistemic component. In Maria Luisa Rivero Ana Arregui & Andrés Salanova (eds.), *Modality across syntactic categories*, 11–30. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198718208.001.0001.
- Alonso-Ovalle, Luis & Justin Royer. 2021. Random choice from likelihood: The case of Chuj (Mayan). *Journal of Semantics* 38(4). 483–529.
- Asis, Jorge. 2000. *Del Flore a Montparnasse: El sentido de la vida en el socialismo*. Buenos Aires: Editorial Sudamericana.
- Baily, Samuel L. 1999. *Immigrants the lands of promise: Italians in Buenos Aires and New York City, 1870–1914*. Ithaca & London: Cornell University Press.
- Boot, Teodoro. 2012. *INTELECTUAL: Adjetivo o sustantivo ¿Qué refalosa es esta palabra!* Pájaro Rojo. <https://pajarorojo.com.ar/?p=3805> (31 July, 2020).
- Brero, Camillo. 2001. *Vocabolario Italiano-Piemontese – Piemontese-Italiano*. Torino: Piemonte in Bancarella.
- Cancellier, Antonella. 2001. Italiano e spagnolo a contatto nel Rio de la Plata: I fenomeni del “cocoliche” e del “lunfardo”. In Antonella Cancellier & Renata Londero (eds.), *Atti del XIX convegno Associazione Ispanisti Italiani, Roma, 16–18 settembre 1999*, vol. 2: Italiano e spagnolo a contatto. Padova: Unipress.
- Carlson, Greg N. 1977. *Reference to kinds in English*. University of Massachusetts. (Doctoral dissertation).
- Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the “logicality” of language. *Linguistic Inquiry* 37(4). 535–590. DOI: 10.1162/ling.2006.37.4.535.
- Chierchia, Gennaro. 2013. *Logic in grammar: Polarity, free choice, and intervention*. Oxford: Oxford University Press.
- Cocco, Maurizio. 2014. *Il Qualunquismo Storico: Le idee, l’organizzazione di partito, il personale politico*. Università degli Studi di Cagliari. (Doctoral dissertation).
- Conde, Oscar. 2011. *Lunfardo: Un estudio sobre el habla popular de los argentinos. (Pensamiento)*. 1st edn. Buenos Aires: Taurus.
- Cortellazzo, Manlio & Paolo Zolli. 1996. *Dizionario etimologico della lingua italiana*, vol. 4. Bologna: Zanichelli.
- Davies, Mark. 2016–. *Corpus del Español*. <http://www.corpusdelespanol.org/web-dial/>.
- De Mauro, Tullio. 2017. *Storia linguistica d’Italia dall’Unità a oggi*. Rome: Editori Laterza.
- Devoto, Fernando. 2002. In Argentina. In Piero Bevilacqua, Andreina De Clementi & Emilio Franzina (eds.), *Storia dell’emigrazione italiana*, vol. 2: Arrivi, 25–54. Rome: Donzelli.

- Di Tullio, Ángela Lucía. 2003. *Políticas Lingüísticas e inmigración*. Buenos Aires: Eudeba.
- Di Tullio, Ángela Lucía. 2015. Cualquiera. In *Diccionario Latinoamericano de la Lengua Española*. Buenos Aires: Universidad Nacional de Tres de Febrero. <http://untref.edu.ar/diccionario/notas-detalles.php?nota=8> (24 February, 2021).
- Diewald, Gabriele. 2011. Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49. 365–390. DOI: 10.1515/ling.2011.011.
- Engels, Kathrin. 2010. *Cocoliche als Mediensprache: Die Darstellung einer Lernervarietät im Theater der Río de la Plata-Raums*. Freiburg: Rombach.
- Engels, Kathrin & Rolf Kailuweit. 2011. Los italo-lunfardismos en el sainete criollo: Consideraciones léxicosemánticas. In Rolf Kailuweit & Ángela Lucía Di Tullio (eds.), *El español rioplatense: Lengua, literaturas, expresiones culturales Iberoamericana* (Lingüística Iberoamericana 51), 227–247. Madrid/Frankfurt: Iberoamericana/Verveurt.
- Ennis, Juan Antonio. 2015. Italian-Spanish contact in early 20th century Argentina. *Journal of Language Contact* 8(1). 112–145. DOI: 10.1163/19552629-00801006.
- Etxeberria, Urtzi & Anastasia Giannakidou. 2014. Anti-specificity and the role of number: The case of Spanish *algún/algunos*. Paper presented at *Rencontres d'Automne de Linguistique formelle : Langage, Langues et Cognition* (RALFe 2014), Université Paris 8, Paris.
- Fălăuș, Anamaria. 2014. (Partially) Free choice of alternatives. *Linguistics and Philosophy* 37(2). 121–173. DOI: 10.1007/s10988-014-9146-8.
- Francia, Marika. 2020. Survey on Italian “qualunque”. https://docs.google.com/spreadsheets/d/1NUxcdi2wbThFbaLpKActyyjIApJClirR/edit?usp=share_link&oid=104417763504507904316&rtpof=true&sd=true (24 February, 2021).
- Geurts, Bart. 2010. Specific indefinites, presupposition and scope. In Rainer Bauerle, Uwe Reyle & Thomas Zimmermann (eds.), *Presuppositions and discourse: Essays offered to Hans Kamp*, 125–158. Leiden: Brill.
- Giannakidou, Anastasia & Josep Quer. 2013. Exhaustive and non-exhaustive variation with free choice and referential vagueness: Evidence from Greek, Catalan, and Spanish. *Lingua* 126. 120–149. DOI: 10.1016/j.lingua.2012.12.005.
- Gutzmann, Daniel. 2013. Expressives and beyond: An introduction to varieties of use-conditional meaning. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning* (Current Research in the Semantics Pragmatics-Interface (CRiSPI) 28), 1–58. Leiden: Brill.
- Haensch, Günther & Reinhold Werner. 2000. *Diccionario del español de Argentina: Español de Argentina - español de España*. 1st edn. Madrid: Gredos.

- Jayez, Jacques & Lucia Tovená. 2007. Subtriggering as alternatives through regularities. In Maria Aloni, Paul Dekker & Floris Roelofsen (eds.), *Proceedings of the sixteenth Amsterdam Colloquium*, 127–132. Amsterdam: ILLC/Department of Philosophy.
- Kailuweit, Rolf. 2004. Spanisch und Italienisch im Spiegel der argentinischen Literatur um 1900: Varietäten- und medientheoretische Überlegungen. *PhiN: Philologie im Netz* 27. 47–66.
- Kellert, Olga. 2021a. *Evaluative meanings of indefinites: A synchronic and diachronic analysis of Spanish cualquiera and related elements*. University of Göttingen. (Habilitation thesis).
- Kellert, Olga. 2021b. Free Choice Indefinites in Old and Modern Italian. *Journal of Historical Syntax* 5(1–13). 1–27. DOI: 10.18148/hs/2021.v5i1-13.82.
- Kellert, Olga. 2021c. The evaluative meaning of the indefinite qualunque in (Old) Italian. In Chiara Gianollo, Klaus von Heusinger & Maria Napoli (eds.), *Determiners and quantifiers: Functions, variation, and change* (Syntax and Semantics 44), 246–284. Leiden: Brill. DOI: 10.1163/9789004473324_009.
- Kellert, Olga. 2021d. *Variation and change of unspecific indefinites: A case study of language contact in Río de la Plata area*. Talk given at LMU on January 1st, 2021.
- Kellert, Olga & Andrés Enrique-Arias. 2024. On the diachrony of Catalan indefinite *qualsevol*. In Olga Kellert, Sebastian Lauschus & Malte Rosemeyer (eds.), *Indefinites in Romance and beyond*, 141–175. Berlin: Language Science Press. DOI: 10.5281/zenodo.13759988.
- Kleparski, Grzegorz A. & Paulina Borkowska. 2007. It befalls words to fall down: Pejoration as a type of semantic change. *Studia Anglica Resoviensia* 47. <https://core.ac.uk/reader/14525410> (30 July, 2020).
- Koch, Peter & Wulf Oesterreicher. 2011. *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Berlin: De Gruyter. DOI: 10.1515/9783110252620.
- Lozza, Arturo Marcos & Osvaldo Pugliese. 1985. *Osvaldo Pugliese al Colón: La vida, la obra y la yumba del maestro del tango, un reportaje exclusivo en tres jornadas y un ensayo orquestal a “rompe y raja”*. Buenos Aires: Editorial Cartago.
- Marchetti, Pablo. 2014. *Puto el que lee: Diccionario argentino de insultos, injurias e improperios*. Puto el que lee: Ediciones Granica.
- Medina, Enrique. 1989. *El secreto*. Buenos Aires: Editores Milton.
- Menéndez-Benito, Paula. 2010. On universal free choice items. *Natural Language Semantics* 18(1). 33–64.
- Meo Zilio, Giovanni. 1959. Sull’elemento italiano nello spagnolo rioplatense. *Lingua nostra* 20. 97–103.

- Meo Zilio, Giovanni. 1964. El “Cocoliche” rioplatense. *Boletín de Filología* 16. 61–119.
- Meo Zilio, Giovanni. 2001. L’elemento linguistico italiano nell’America Latina: Altri italianismi del rioplatense nel repertorio di Teruggi. *Rassegna Iberistica* 73. 51–59.
- Piccitto, Giorgio & Giovanni Tropea. 1990. *Vocabolario siciliano*, 3 vols. Palermo: Centro di studio filologici e linguistici siciliani.
- Quer, Josep. 2000. Licensing free choice items in hostile environments: The role of aspect and mood. *SKY Journal of Linguistics* 13. 251–268.
- Real Academia Española. 2010. Cualunque. In *Diccionario de americanismos*. Madrid: Santillana. <http://lema.rae.es/damer/?key=cualunque> (30 July, 2020).
- Rizzo Salierno, Melisa. 2013. El uso adjetivo de cualquiera. In Ángela Di Tullio (ed.), *Aproximaciones al estudio del español de la Argentina*, 23–31. Buenos Aires: Universidad Nacional del Comahue.
- Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.
- Romano, Roberto. 1999. *El tren de las diez*. Buenos Aires: Ediciones Simurg.
- Seiler, Guido, Dankmar Enke & Roland Mühlenbernd. 2018. *Stability and instability in grammar*. Presented at the 20th International Congress of Linguistics, Cape Town.
- Silberstein, Enrique. 1967. *Charlas económicas*. Buenos Aires: A. Peña Lillo.
- Stark, Elisabeth. 2006. *Indefinitheit und Textkohärenz: Entstehung und semantische Strukturierung indefiniter Nominaldetermination im Altitalienischen*. Tübingen: Max Niemeyer Verlag.
- Stein, Dieter. 1990. *The semantics of syntactic change: Aspects of the evolution of “do” in English*. Berlin: De Gruyter. DOI: 10.1515/9783110846829.
- Teruggi, M. 1974. *Panorama del lunfardo*. Buenos Aires: Ediciones Cabargón.
- Traugott, Elizabeth C. 1989. On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language* 65(1). 31–55. DOI: 10.2307/414841.
- von Heusinger, Klaus. 2011. Specificity. In Klaus von Heusinger, Claudia Meinborn & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 2, 1024–1057. Berlin: de Gruyter.
- Zamparelli, Roberto. 2000. *Layers in the determiner phrase*. University of Rochester. (Doctoral dissertation).

Chapter 4

Indefinite pronouns with **THING** and **PERSON** in two Ibero-Romance/Kikongo varieties: Palenquero Creole and Cabindan Portuguese

Miguel Gutiérrez Maté

University of Augsburg

This chapter deals with the origins of generic-noun-based indefinites (according to the terminology of the WALS) in two Ibero-Romance varieties, the Spanish-lexified Palenquero Creole and the heterogeneous group of postcolonial varieties of Portuguese that are spoken in the exclave of Cabinda (Angola). Both varieties have in common the fact that they have been influenced by the same “substrate”, the Bantu H language Kikongo. Both substratal influence and language universals during first- and/or second-language acquisition seem to interact in the making of indefinite expressions, as they always interact in restructuring phenomena found in language contact ecologies. However, as regards indefinite expressions, there are substantial differences between Palenquero and Cabindan Portuguese, due to the fact that the latter often retains the special indefinites from the superstrate, whereas most Spanish special indefinites have been lost in the former (with some exceptions, most especially *ná* < *nada*). The two varieties studied here result from different language contact ecologies, which account for quantitative and qualitative differences between the two varieties.

1 Introduction

During my fieldwork in the village of San Basilio de Palenque (Bolívar, Colombia) in the summer of 2017, a traditional speaker of Palenquero Creole (RC, 84 years



Miguel Gutiérrez Maté. 2024. Indefinite pronouns with **THING** and **PERSON** in two Ibero-Romance/Kikongo varieties: Palenquero Creole and Cabindan Portuguese. In Olga Kellert, Sebastian Lauschus & Malte Rosemeyer (eds.), *Indefinites in Romance and beyond*, 87–139. Berlin: Language Science Press.
DOI: 10.5281/zenodo.13759986

old, female) without being asked started to teach me some good old-fashioned Palenquero¹ (or just *Traditional Palenquero*, as opposed to the Palenquero varieties spoken by adult heritage speakers and the learner varieties that children learn at the local school; see Lipski 2020 for an in-depth exploration of the cohabitation of all these varieties in the village alongside local Spanish). She corrected my Palenquero (henceforth PAL) on some occasions, as in the following example:

RC: utere tá kombesá?

you(PL) PROG talk

‘are you talking?’

MG: bueno... un poko

well... a little

‘well... a little’

RC: un poko kusa!

a little thing

‘a little’

MG: un poko kusa

a little thing

‘a little’

My mistake consisted in the use of *un poko* in the exact same way as its Spanish source *un poco*, i.e. as a well-formed indefinite expression – which, by the way, is used adverbially in this context. When transferred into PAL, however, the structure is ill-formed: Sp. *conversar un poco* → Pal. **kombesá un poko*. As

¹In today’s San Basilio de Palenque, a village where the visits of scholars and students (or just tourists) have become quite regular in the last years, it would not be impossible that some informants, when asked for translation of a given Spanish sentence into PAL or for the correction of PAL sentences, would even make up their Palenquero in order to make it lexically and/or structurally more distant from Spanish, i.e. more “exotic”. However, this was clearly not the case here: first, the fragment above is taken from the second interview I made with this speaker, after our having developed some mutual trust the day before, which turned out to be relatively easy since I was accompanied by a youngster from the village whom she knew well; second, the conversation was not dealing with linguistic issues (it was just about her life and the history of the part of the village she was living in) and she was not even asked to teach us PAL: she rather invited us to bring the conversation to a new level, after we spoke in Spanish and she spoke in PAL for a while; third, and perhaps most importantly, she was not expecting any money for the interview and she never actually asked for it (in fact she seemed to enjoy the company quite a lot).

the correction introduced by this particular informant clearly pointed out, indefinites require an explicit nominal head in PAL: Thus, when the indefinite is to be understood as a pronominal, not modifying any noun, the head has to be occupied by the generic nouns *kusa* ‘thing’ (< Sp. *cosa*), for inanimate expressions (*aggú kusa* ‘something’ [lit. ‘some-thing’]), or *hende* ‘people’ (< Sp. *gente*), for personal expressions (*aggú hende* ‘someone/somebody’ [lit. ‘some people’]). In fact, an anecdote of this kind makes me think that the degree of certainty with which the feature 21 of the *Atlas of Pidgin and Creole Language Structures* (APICS) (“indefinite pronouns”) is codified as generic-noun-based in PAL should be changed from “certain” (as prudently indicated by Schwegler 2013) to “very certain”. It is true, however, that, due to the long-term bilingualism PAL/Spanish in the village, we can expect special indefinites like *aggo* ‘something’ (< Sp. *algo*) or *agguno* ‘someone’ (< Sp. *alguno*) to be incorporated in some varieties of PAL, especially in those by speakers whose bilingualism tends, for some reason or another, towards Spanish.

The above anecdote is meant to give a first idea of what the following pages are about. My aim in this chapter is to examine the distribution and origins of such generic-noun-based indefinites in PAL and, in doing so, reflect on the genesis of these kinds of indefinites in Creole languages, i.e. on its role during the process of creolization. To this end, the comparison of PAL indefinite expressions with those that are found in restructured Cabindan Portuguese (henceforth, CP) – a complex of postcolonial varieties of Portuguese spoken alongside Kikongo in the exclave and province of Cabinda (Angola) – will reveal itself as extraordinarily helpful as a means of understanding the limits between different types of language contact varieties (which includes setting quantitative and/or qualitative limits between creolization and second language acquisition processes). This is so due to the fact that both PAL and CP came into being when speakers of Kikongo varieties² somewhat “approximated” an Ibero-Romance language (Spanish in the case of PAL, Portuguese in the case of CP), which led to more or less intense restructuring of Ibero-Romance: see Holm (2004) about *partial restructuring* and Holm et al. (1999) about different degrees of restructuring, where Creoles are considered to be one end of the “restructuring continuum”.

The term *language approximations* (Chaudenson 2003) is often preferred to *learner varieties*, since it would not be accurate to state that Creoles (like PAL) resulted from canonical language learning: First generations of Creole speakers

²Bantu H10-16 according to Guthrie’s most famous classification of Bantu languages (Guthrie 1967–1971), *Kikongo Language Cluster* according to Bostoen (2012) and Bostoen & Schryver (2015).

did not really try to *learn* another language in the modern sense of the term, but they just wanted to be able to communicate with other speakers with whom they did not share, at first, a common language. In doing so, it became clear, however, that speakers had – partially, at least – to give up their own first language in their effort to adapt to the primitive Creole societies (plantations with slaves from different African regions, maroon communities formed out of the reunion of bozals – native speakers of African languages – and black Creoles – native speakers of regional varieties of European languages, etc.); it is only in the latter sense that we can relate creolization to second-language (L2) acquisition.

If we extend the model of Michaelis (2017) – originally designed for the classification of Creole languages according to their lexifiers and substrates – to any other vernacular resulting from a language contact situation, we should classify both PAL and CP as Ibero-Romance/Bantu or, more precisely, Ibero-Romance/Kikongo. Moreover, if we wanted to highlight the role of a particular Kikongo dialect in the formation of PAL we could possibly speak of a Spanish/Kiyombe Creole, since it is surely Kiyombe – the autoctonous variety of the Mayombe forest, from which many Black slaves were taken and brought to the New World – that constitutes PAL's main substrate (s. Moñino 2017, Schwegler 2016a, 2017, Gutiérrez Maté 2020 and references therein).³ To ensure comparability, unless indicated otherwise, the Cabindan examples selected for this chapter have been produced by Portuguese/Kiyombe bilinguals, who I interviewed in the Cabindan Mayombe (municípios of Buco Zau and Belize). However, I do not expect there to be substantial differences between the Kikongo varieties (certainly not between the Cabindan varieties of Kikongo) as regards the particular phenomenon analyzed throughout this chapter.

³It has to be noted that Moñino (2017: 24–25) states that Vili (spoken along the Loango coast) was the language of the regional slave trade during the 17th century and could therefore have become “the primary base of the Congo substrate of Palenquero”. Speakers of Kiyombe and other varieties could have easily learned Vili as an L2 or just have adapted to it (i.e. Vili might have served as the basis for a Kikongo koine). The hypothesis is suggestive, but – as Moñino himself acknowledges – we only have evidence of the genetic match between the inhabitants of Palenque and those from the Mayombe forest (no match between Palenqueros and the people from the Loango coast has been shown to date). Consequently, I assume that the Yombe people were the most important group in the primitive Palenquero society and that (L1-)Kiyombe had as good “adaptive” chances as (L2-)Vili in the New World setting. In reality, however, we can trace the origin of some PAL features back to Kiyombe and the origin of some other features back to several westernmost varieties of Kikongo (including Vili) (Gutiérrez Maté in preparation). When it comes to CP, the ultimate reason for focusing on Kiyombe-Portuguese bilinguals in my study is the fact that there are not so many Vili-Portuguese bilinguals in Cabinda, since Vili is only spoken in a small region bordering the Republic of the Congo.

The orientation of my work is mainly typological, with focus on the emergence of new languages/varieties out of the contact between languages belonging to different types. My understanding of “generic-noun-based” indefinites follows the tradition of the *World Atlas of Language Structures Online* (WALS, Haspelmath 2013), which, for its part, follows the work of Haspelmath (1997). According to this author, indefinite pronouns are often derived from “generic ontological-category nouns” such as THING, PERSON, PLACE, MANNER, etc. (Haspelmath 1997: 26, 52). In many cases, the pronominal status of such expressions is controversial, and a certain degree of idealism is needed:

for most languages with generic-noun-based indefinites, there is no good evidence available that these expressions are different from ordinary indefinite noun phrases. In fact, some descriptions explicitly deny that they are indefinite pronouns. [...] In this chapter, such cases where we lack evidence for pronominal status have been lumped together with languages like Italian and English because it would have been very difficult to draw a line between the two types. The evidence for pronominal status can be rather subtle. For instance, French *quelque chose* ‘something’ at first glance looks like a regular indefinite noun phrase consisting of the indefinite determiner *quelque* and the noun *chose* ‘thing’. However, the notion ‘something good’ is expressed with a construction that is reserved for pronouns: *quelque chose de bon* (cf. *quoi de bon* ‘what good’), whereas an ordinary noun phrase would be *quelque bonne chose* (‘some good thing’). (Haspelmath 2013: §2.2)

At the very least, we could state that a given language exhibits “generic-noun-based indefinites” when it uses NPs with ontological-category nouns in the phrasal head to convey the meaning of “what other languages express by means of indefinite pronouns” (Haspelmath 1997: 28). We could assume that, prototypically, languages classified as having generic-noun-based indefinites should only exhibit generic-noun-based pronominals; however, it may be convenient to include other languages in the same group even if they have generic-noun-based indefinite expressions that 1) are not clearly grammaticalized (nor lexicalized) as such pronouns and/or 2) have not completely replaced the paradigm of special indefinites.

Let us consider, for example, the case of European Portuguese: in this language, the most common indefinite expression meaning ‘something’ is *alguma coisa* (‘some thing’), which is obviously generic-noun-based; however, unlike English *something*, it does not actually seem to have undergone pronominalization, nor has it completely replaced the special indefinite *algo*. Despite all this, the fact

that an NP projected by a general ontological-category noun has become the indefinite expression *par excellence* for conveying the meaning of ‘something’ should be factored into our typological characterization of Portuguese. This is also the reason why the WALS eventually classifies Portuguese as a language having “mixed indefinites” (Haspelmath 2013: §2.4); the latter means that, although Portuguese makes use of a special indefinite for the ‘someone’ word (*alguém*), it does frequently exhibit a generic-noun-based expression (*alguma coisa*) for the meaning of ‘something’.

In other words, it is not only the categorical use of a given linguistic strategy, but also its productivity, that matters for linguistic typology. In fact, throughout my work, most especially as regards CP, I will be dealing with generic-noun-based indefinite expressions of which all I can say with certainty is that they are highly productive (cases like the above, in which a native speaker explicitly rejects an alternative strategy, are generally not found).

The structure of this chapter is as follows: In §2, the PAL noun-based indefinites exemplified in this introduction will be presented in more detail, including a brief description of other uses of generic nouns, in order to have a general idea of their semantic heterogeneity and the ease with which they can undergo semantic bleaching and pronominalization. §3 is about investigating the role of three aspects – which are all essential parts of any Creole – in the formation of indefinites in PAL: (a) linguistic universals during first and/or second language acquisition, (b) further development of tendencies that already existed in the lexifier (or superstrate) and (c) substratal influence. To further illustrate the latter (which I consider to be *conditio sine qua non* in the process of change analyzed here), indefinite pronominal expressions in CP will be examined in §4. The chapter ends with a summary of the main results and their contextualization within the framework of a wider research program (§5).

The data of CP and Kikongo were collected *in situ* as a result of my fieldwork in Cabinda in March/April 2019 and February/March 2020, whereas the data from PAL result from combining my own data (from fieldwork made in 2017) with those from the corpus published by de Friedemann & Rosselli (1983) and Maglia & Moñino (2015) and those from the first interviews made by A. Schwegler in San Basilio de Palenque (1985–1988).⁴ The corresponding source is indicated after each PAL example.

⁴ A corpus of recordings I have been transcribing, digitizing, and analyzing since 2014 (cf. Gutiérrez Maté 2017, 2020).

2 The Palenquero data

2.1 *kusa* ('thing') and *hende* ('people') in indefinite expressions

As stated above, indefinites in PAL are mainly built by using generic nouns: this is not only the case for indefinites meaning 'something' and 'someone', which are those that both the WALS and the APICS take into account for their typological classifications, but also for indefinites meaning 'everything', 'a lot', 'a few', etc. (all of them having in common the fact that they are quantificational expressions, be they "universal" or not, "evaluative" or not, etc.: Bosque & Gutiérrez-Rexach 2009: 501–502). When the indefinite is positive and non-personal, a determiner is always expressed and placed preminally. A few examples of different types of generic-noun-based indefinites will suffice to demonstrate this trend in PAL. Particularly, as for the expression of 'something', PAL exhibits two forms, since the selection of the determiner seems to almost freely alternate between *un* and *aggú* (we cannot disregard the possibility that each form has different semantic nuances or is even conditioned by different syntactic constraints, but the available data do not allow any conclusions to be drawn yet):

- (1) *aggú/un kusa* 'something' (= Sp. *algo*):
 - a. (Schwegler 2013: 48–33)
 Bo a komblá aggú kusa?
 2P.SG CPL buy some thing
 'Did you buy something?' (Sp. *¿Compraste algo?*)
 - b. (Maglia & Moñino 2015: 261)
 <antonse vamo a hacía> un kusa pa nu José akkansá suto
 <thus we-are-going to do> a thing for NEG José reach us
 'then let's do something so that José cannot reach us'
 (Sp. *Entonces vamos a hacer algo para que José no nos alcance*)
- (2) *to kusa* 'everything' (= Sp. *todo*):
 - a. (Recordings by Armin Schwegler 1985–1988)
 Aora jue la mora tá ke hende a ten ke asé to kusa pa moná
 Today FP the way be that people CPL have to do all thing for child
 chikito
 small
 'Nowadays, there's a trend that people have to do everything for children' (Sp. *Ahora está la moda de que la gente tiene que hacer todo para los niños*)

- b. (Recordings by Armin Schwegler 1985–1988)
 y'ase-[b]a hundá-lo to kusa
 1P.SG. HAB-IMP gather-3P.OBJ all thing
 'I used to gather everything together' (Sp. *solía juntarlo todo*)
- (3) *mucho kusa* 'much/a lot' (= Sp. *mucho*):
- a. (Recordings by A. Schwegler 1985–1988)
 bo <biene> má pokke bo etá yebá mucho kusa
 2P.SG <come> more because 2P.SG PROG bring much thing
 'you will come back more often, because you are gaining (/learning) a lot from here' (Sp. *Vendrás más porque te estás llevando[/estás aprendiendo] mucho [de aquí]*)
- b. (de Friedemann & Rosselli 1983: 215)
 kumo suto ten kampo nu, akí ta pelé mucho kusa
 due.to 1P.PL have field NEG, here PROG lose much thing
 'since we don't have fields, a lot is being lost here' (Sp. *Como no tenemos campos, aquí se está perdiendo mucho*)

A different case is represented by the use of *kusa* in negative polarity contexts. In these contexts, *kusa* is generally employed without any indefinite determiner:

- (4) (Maglia & Moñino 2015: 214)
 aora ma hende ta ase-ndo ele <en la noche> [...] pa nu hende
 now PL people PROG do-GER 3P.OBJ [at night] for NEG people
 nu ndá kusa⁵
 NEG give thing
 'now, people are doing it at night [...] so as not to give anything'
 (Sp. *Ahora la gente está haciéndolo de noche [casarse] para no dar nada*)

⁵Final sentences are quite exceptional in PAL, since they use only one negator, which is placed before the subject and immediately after *pa* (< Sp. *para*), as in the example (1b) above. Consequently, in (4), we would have expected *pa nu hende ndá kusa*. All other sentence types, including matrix sentences, have three possibilities to convey negation: with a preverbal negator, which is always placed *after* the subject (NEG1), with both a preverbal and a sentence-final negator (NEG2) and with a sentence-final negator (NEG3) (Schwegler 2016b). The additional use of the negator *nu* between the subject and the verb might therefore have been influenced by two of the three regular negation types (NEG1 and NEG2). On the other hand, the speaker might just be producing a spontaneous mixture of a Palenquero final sentence and a canonical Spanish sentence (with NEG1). Since final sentences have mostly remained unexplored in the literature on PAL negation, it is hard to tell whether this particular type of "preverbal double negation" is more common or even regular in some idiolects. From a purely structural view-

- (5) (Maglia & Moñino 2015: 242)

Suto polé-ba asé kusa malo ante nu
1P.PL can-IMP do thing bad before NEG

‘we could not do anything wrong before.’ (Sp. *No podíamos hacer nada malo antes*)

It has to be noted that PAL also makes use of the special negative indefinite *naa* or *ná* (< Sp. *nada*). Unlike other Spanish-like special indefinites, *ná* seems to have been fully incorporated in the Creole (Schwegler 2016b: 234) and traditional speakers use it regularly (in some contexts, alongside the generic-noun-based form). The factors accounting for the variable of negative non-personal indefinites (“*ná* vs. *kusa*”) in negative polarity contexts are yet to be determined.

As regards *hende*, we find a similar distribution to the one observed in the case of *kusa*. The determiner decides the type of quantifying reading (in this particular example, we have the universal one: *to hende* = ‘everyone’):

- (6) (Fieldwork M. Gutiérrez Maté 2017)

ma hende <preguntando> pu[sic] to hende: Raú kiene jue?
PL people asking for all people: Raúl who is?

‘people are asking everyone: “Who is Raúl?”’

(Sp. *La gente [anda] preguntando a todos/todo el mundo: “Raúl quién es?”*)

Special indefinites like *nadie* ~ *narie* (< Sp. *nadie*) and *aggie* (< Sp. *alguien*) are nearly non-existent in the traditional PAL varieties analyzed here. However, we have to deal with the fact that special indefinites with a partitive reading seem to be fully integrated in the Creole: *aggú(n)* (< Sp. *algún*), *agguno* (< Sp. *alguno*) and *ninguno* (< Sp. *ninguno*). The latter can also be used as an inherently negative word with the meaning of ‘nobody’/‘no one’.⁶

point, I guess that scholars working on the generativist framework would treat this example as one of those exceptional cases in which a moved element (here, the negator) receives a phonological representation in the different structural positions that it has during the derivation: see the analysis of sentences like *Wen meinst du wen Peter gewählt hat?* in German dialects by Gabriel et al. (2018: 114).

⁶Let us consider the following example:

- (i) (Recordings by A. Schwegler, 1985–1988)

ninguno sa[b]é ké kusa é bitibite nu
no.one know what thing COP bitebite NEG

‘no one knows what bitebite [a traditional food] is’

In addition, we have to deal with the peculiarity that, in some cases, we cannot decide whether we have an indefinite pronoun or a generic/arbitrary expression (similar to the Spanish source *gente*). This is also a hint about the blurred limits between indefinite and generic expressions in many contexts (cf. Sp. *No ha venido gente/No ha venido nadie*) (I will come back to these issues in §2.2). The examples illustrate the use of *hende* as a negative indefinite (without determiner):

- (7) (Maglia & Moñino 2015: 287)
 Uto pueblo bo miná hende asina nu
 other village 2P.SG look someone/people like.this NEG
 ‘In other villages you don’t see anyone/people like that’
 (Sp. *En otros pueblos no ves nadie así/gente así*)
- (8) (Maglia & Moñino 2015: 288)
 Aki Palenge a ten kumina po lendro monte [...] ke ma
 here Palenque CPL EXIST food over inside mountain that PL
 hende kelé-lo nu
 people want-it NEG
 ‘Here in Palenque there is food in the forest that [...] nobody wants to [go and get]’ (Sp. *Aquí en Palenque hay comida en el monte que nadie quiere [ir a buscar]*)

As (8) shows, the pluralizer *ma* often precedes *hende*: actually, *hende* and *ma hende* alternate quite freely, a fact that might be related to the history and meaning of *ma*. The source for this PAL item is the noun class 6 prefix in Kikongo (Schwegler 2007, Moñino 2013); in this language (as well as in Proto-Bantu), *ma-* is a productive plural prefix⁷ and also the prefix used for liquids, masses and collectives (see Chicuna 2018: 108–109 for Kiyombe and Maho 1999: 51 for Proto-Bantu). Since PAL does not distinguish noun classes morphologically, *ma* has become the only pluralizer in the Creole, or, better said, an optional pluralizer, due to the fact that plural can also be interpreted contextually, with no need of marking it morphologically (Moñino 2013: 42–43). Interestingly, in PAL *ma-* has been lost in mass nouns (*agua* ‘water’, *asuka* ‘sugar’, etc.) (Moñino 2013: 56–57), but it can still be used for collectives: for instance, *ma ngombe* (where *ngombe* ‘cow’ is another Bantuism in PAL) can sometimes be better translated as ‘cattle’ than ‘cows’. Consequently, the alternation “*hende* vs. *ma hende*” can be seen as the result of two different issues that imply opposing tendencies in PAL: on

⁷Not only is *ma-* the plural of class 5 (*ditoko/matoko* ‘boy/s’) but also the plural of classes 14 (*bwala/maala* ‘village/s’) and 15 (*kulu/malu* ‘leg/s’).

the one hand, *ma-* is mostly specialized as a plural marker but it can also retain other etymological meanings (including the collective); on the other hand, speakers tend to use *ma* for conveying the aforementioned meanings but, if these can also be understood contextually, there is no need to use *ma* at all.

Table 1 sums up the indefinite expressions that are found in PAL.

Table 1: Generic-noun-based indefinites in PAL

[−personal]	[affirmative]	<i>un~aggú kusa</i>	‘something’
		<i>to kusa</i>	‘everything’
		<i>mucho kusa</i>	‘a lot’
		<i>un poko kusa</i>	‘a few, a little’
	[negative]	<i>ná</i>	‘nothing’
		(NEG+) <i>kusa</i>	‘(not...) anything’
		~(NEG+) <i>ná</i> (less frequent)	
[+personal]	[affirmative]	<i>un~aggú hende</i>	‘somebody/someone’
		<i>to hende</i>	‘everyone’
		<i>mucho hende</i>	‘a lot of people’
		<i>un poko hende</i>	‘a few people’
	[negative]	<i>ninguno</i>	‘nobody/no one’
		(NEG+) <i>hende~ma hende</i>	‘no one/(not...) anyone’

2.2 Other uses of *kusa* and *hende*

Even though the nouns *kusa* ‘thing’ and *hende* ‘person/people’ are frequently used as indefinite pronominals in PAL, we have to acknowledge the fact that they have a wide spectrum of uses, some of which are more lexical and more prototypically nominal than others. In my view, one can easily admit that changes of the type ‘a/some thing’ > ‘something’ involve grammaticalization (as perhaps in the case of *un/aggú kusa* and similar indefinite expressions); as for those cases in which the generic noun has no determiner in negative contexts (as when *kusa* means ‘anything’), we could analyze them as a particular case of semantic bleaching, i.e. in a similar fashion as the previous literature has analyzed the change Lat. REM > Old French *rien(s)* > French *rien* (s. Roberts 2012: 364). As in any other linguistic change, the innovative meaning did not suddenly replace (and may have never completely replaced) the traditional one (Hopper 1991). In this section, I give a brief account of the uses in which *kusa* and *hende* are not to be

understood as indefinite pronominals (which is not to say that all these uses are necessarily conservative). To begin with, *kusa* is prototypically nominal when it refers to (specified or unspecified) material things (as in (9), but not in (10)):

- (9) (Maglia & Moñino 2015: 247)
 ma loke nu ten moná ju' i ta jutá kusa akí
 ART REL NEG have child COP REL PROG steal thing here
 'those who have no children are those who are stealing things here'
 (Sp. *Los que no tienen hijos son los que están robando cosas aquí*)
- (10) (Recordings by A. Schwegler 1985–1988)
 entonse kusa a ñamá mí <la atención> pokke y' a mina
 thus thing CPL call 1P.SG the attention because 1P.SG CPL see
 kúmo to ma pueblo <tenía[n]> karretera ané
 how all PL village had road 3PL.POSS
 'So one thing caught my attention because I saw that all the villages
 [around here] [already] had their roads' (about cars getting stuck at the
 entrance of the unpaved road leading to San Basilio de Palenque)
 (Sp. *Entonces eso [este hecho] me llamó la atención porque yo vi cómo todos
 los pueblos tenían ya sus carreteras*)

In a similar vein, *hende* is a canonical noun when it refers to some (specified or unspecified) person. As a matter of fact, the Spanish noun *persona* did not really get into PAL, so *hende* (< *gente*) unites both the meaning of the Spanish generic noun *gente* 'people' and the more concrete *persona* 'person':⁸

- (11) (Fieldwork M. Gutiérrez Maté 2017)
 to e[se] ma hende i a konosé-lo nu
 all that PL people 1P.SG CPL know-3P.OBJ NEG
 'all those people [you have just mentioned] I do not know'
 (Sp. *A toda esa gente [/a todas esas personas] no las conozco*)
- (12) (Recordings by A. Schwegler 1985–1988)
 entonse a tene-ba ndo hende nu-má'kí
 then CPL EXIST.have-IMP two people no-more'here
 'back then there were just two people here'
 (Sp. *Entonces solo había dos personas aquí*)

⁸As one of the reviewers pointed out, Spanish speakers often use sentences in which the distinction generic/concrete is blurred: for example, *Juan es muy buena gente* (with the literal meaning 'Juan is very good people' and the actual meaning 'Juan is such a nice guy').

Many uses of both *kusa* and *hende* (like those preceded by the indefinite determiner) seem to already find themselves on the limit between nominal and pronominal expressions: ‘a (given) thing/something’, ‘a (given) person/someone’ (the same is actually true for noun phrases with ‘thing’ and ‘person’ in Ibero-Romance – and many other languages – when they have a non-specific reading, i.e. when the potential referents are interchangeable):

- (13) (Recordings by A. Schwegler 1985–1988)
pero aora bo temé di betí un kusa aí kueppo sí
but now 2P.SG be.afraid of wear a thing there body 2P.SG.POSS
pokke bo polé bití-lo nu
because 2P.SG can wear-3P.OBJ NEG
‘but now you are afraid to put something on your body, because you cannot wear it’
(Sp. *Pero ahora tienes miedo de ponerte una/cualquier cosa[/algo] en tu cuerpo, porque no puedes vestirlo*)
- (14) (Maglia & Moñino 2015: 209)
si un hende andi mitá kaya asé-ba hablá suto-ba [...]
COMP(COND) a person where middle street HAB-IMP tell 1P.PL-IMP
‘if a person(/someone) in the middle of the street spoke to us, [...]’
(Sp. *Si alguien en medio de la calle nos hablaba, [...]*)

As regards *kusa*, we can easily observe other secondary uses, which can all be considered to be derived from the non-material meaning of the generic noun (see 10). These other uses of THING can be very diverse cross-linguistically and adopt different discursive and informational values (I am thinking, for example, of uses like Eng. *Thing is...*, German *Hauptsache...*, etc., which highlight the utterance they are introducing and somehow contrast it with what has been previously said). It is not strange that such uses acquire connective properties which can be equally diverse. For instance, in (15) it can function as an inter-sentential connector (comparative or consecutive: ‘so that, in such a way’):

- (15) (de Friedemann & Rosselli 1983: 203)
kuando bo kabá ese punchera, bo a rregresá, gobbí yená
when 2P.SG finish that basin 2P.SG CPL come.back come fill
punchera má pa gobbí salí, kusa kuando Tito ke paresé ri
basin more for again leave so.that when Tito VIRT appear from
á Katajena, suto a tá lito
there Cartagena 1P.PL COMP be ready
‘when you are finished with this bowl, you return and fill it again in order

for you to (be able to) leave, so that when Tito comes back from there, from Cartagena, we are ready (to leave)'

(Sp. *Cuando uno ha acabado esa "ponchera", uno regresa y vuelve a llenar la "ponchera" otra vez. Para volver a salir, de manera que cuando Tito aparezca de allá de Cartagena, nosotras estaremos listas*)

Besides the change of *kusa* from a generic noun to an indefinite pronoun we also find the change to an interrogative pronoun (introduced by *ké* < Sp. *qué*): in other words, *ké kusa* (lit. 'what thing') is a common variant of *ké* ('what') (actually, what we usually find in these cases is the interrogative pronoun followed by the focus particle *jue* [hwe]: *ké (kusa) jue bo tá asé?* 'what are you doing?'; cf. Gutiérrez Maté 2017). However, generic-noun-based interrogative pronouns with the meaning of 'who?' do not seem to be allowed in PAL: in other words, the variant *ké hende?* ('what people') has never been found, neither in my corpus nor in the other PAL corpora, and it is only the special interrogative *kiene* (< Sp. *quién*) that can be used for this meaning.

As for other uses of *hende*, it can also expand semantically and adopt other meanings, including that of a *generic* or *arbitrary* subject pronoun (the latter being different from the former insofar as its reference explicitly excludes the speaker: cf. Holmberg et al. 2009: 63–64): going one step further, *hende* can even function as a sort of 1P.PL pronoun (cf. Schwegler 1993, 2002). In (16) *hende* is used as something between a generic/arbitrary noun and a 1P.PL pronoun, whereas in (17) it is clearly used as a 1P.PL possessive (possessives in PAL consist of independent personal pronouns placed in a post-nominal position). Consequently, the type of linguistic change we are dealing with parallels the one that has been taking place in Brazilian Portuguese from the 19th century onwards (Lopes 2003) – the grammaticalization degree of *gente* in Brazilian Portuguese⁹ being much more advanced than that of *hende* in PAL:

- (16) (Recordings by A. Schwegler 1985–1988)
asina jue hende asé abla-lo-ba akí... asina jue-ba
so COP/FP people HAB say-3P.ACC-IMP here so COP/FP-IMP
'that's how people used to call it (/how it was called) here... so it was'
(Sp. *Así es como la gente solía [/nosotros solíamos] llamarlo aquí... así era*)

⁹Even in some central and southern varieties of European Portuguese, as one of the reviewers pointed out.

- (17) (Recordings by A. Schwegler 1985–1988)

bo etá bibí <con> un mujé asina kumo koló hende?
 2P.SG PROG live <with> IND.ART woman like.this as colour 1P.PL.POSS
 ‘are you living with a woman like that, who has the same color as us
 [=with a black woman]?’
 (Sp. *¿Estás viviendo con una mujer así, que tiene el mismo color que nosotros?*)

All the above data is interesting because it shows that there are different processes of grammaticalization with regard to generic nouns taking place at the same time; in addition, this line of reasoning leaves the door open for the possibility that the change “generic noun > indefinite expression” (as any other linguistic change involving generic nouns) has occurred within the diachrony of PAL itself. According to such explanation, we would say, in traditional terms, that generic-noun-based indefinites result from an “internal” linguistic change. Be that as it may, there are other uses of *kusa* and *hende* in PAL that seem to be formally and/or semantically related to their use in indefinite expressions, even though we cannot know whether such a relation resulted from a chain of internal changes in PAL internal linguistic history or took place more or less at the beginning (during creolization). I particularly wish to highlight the structural resemblance of the generic-based-noun indefinites to those generic nouns that appear in phrases in which, in the lexifier language, there would be no overt generic noun but only an adjective, a free (or *headless*) relative clause, etc. In other words, we expect many more nominal heads (preceded by determiner or not) to be filled up with a generic noun in PAL than in Spanish (this kind of construction is possible in PAL, but the use of *kusa* and *hende* in this context seems to be far more frequent than the use of, respectively, *cosa* and *gente* in Spanish in the same structural contexts):

- (18) (Maglia & Moñino 2015: 247)

Ndá ri kuenda un kusa lok’i tan ablá bo aora
 Give of notice one thing REL-1P.SG(CL) FUT tell 2P.SG now
 ?Ndá ri kuenda Ø lok’i tan...
 ‘Realize one thing that I am going to tell you now’/‘Realize what I am going to tell you’
 (Sp. *Date cuenta de lo que te voy a decir ahora*)

- (19) (Fieldwork M. Gutiérrez Maté 2017)

Ele é prieto... [ele é] hende... hende kumo suto
 3P.SG COP black [3P.SG COP] people people like 1P.PL
 ?Ele é Ø kumo suto

‘He is black... he is like us’ (an old woman speaking about Armin
 Schwegler, who is well known in the village)
 (Sp. *Él es negro [en realidad], es como nosotros*)

Finally, some combinations of generic noun and adjectives seem to even have lexicalized, as when *kusa* is modified by *di belá belá* (lit. ‘of true true’) (*kusa di belá belá* = ‘the actual/real truth’) or when *hende* is modified by *ngande* ‘big’ (*hende ngande* ‘adult(s)’).¹⁰

- (20) (Recordings by A. Schwegler 1985–1988)

yo ju’i te¹¹ ablá té belá, yo ju’i te ablá
 1P.SG FP-1P.SG(CL) PROG tell 2P.SG truth, 1P.SG FP-1P.SG(CL) PROG tell
 té kusa di belá belá
 2P.SG thing of truth truth

‘I am the one who is telling you the truth, I am the [only] one who is
 telling you the actual truth’

(Sp. *Yo soy el que te dice la verdad, yo soy el que realmente te está diciendo la verdad*)

¹⁰ *Di belá belá* can be combined with other nouns (f.i., *amigo di belá belá* ‘a true friend’: cf. Schwegler 1996: xvii), but *kusa di belá belá* seems to be the lexical expression of ‘(real) truth’ (more than a simple truth, since in the latter case *belá* would surely have been enough).

¹¹ According to the context, the best way of interpreting this sentence is with a progressive verb tense. Such interpretation makes us think that preverbal *te* is a variant of the TMA marker *ta* (progressive), even though I am not sure whether I have even found such a variant before (neither in Schwegler’s oldest recordings nor in any other corpus). The form might result from (regressive) assimilation to the front vowel of postverbal 2P object *té* (< Sp. *usted*). Another possible interpretation would consist in thinking of *te* as the result of joining the focus particle *é* to the TMA *tá* (*tá* + *é* = *té*), but, in PAL, the focus particle cannot be placed between the TMA and the verb (in Colombian Spanish, however, the sequence “auxiliary verb + focus particle/(“focalizing ser”) + main verb” is very common: *estoy es diciéndote la verdad*) (Gutiérrez Maté 2017: 18). Finally we could consider *te* to be the Spanish 2P clitic pronoun, inserted in a sentence that, for the most part, is constructed in PAL. According to this reading, the analysis would be:

(i) yo ju’i <te> ablá té belá, yo ju’i <te> ablá té
 1P.SG FP-1P.SG(CL) <2P.CL>-PROG tell 2P.SG truth, 1P.SG FP-1P.SG(CL) <2P.CL> tell 2P.SG
 kusa di belá belá
 thing of truth truth
 ‘I am the one who tells you the truth, I am the one who tells you the actual truth’

- (21) (Recordings by A. Schwegler 1985–1988)
 to majanasito chikito a tá kamino ané i hende ngande
 all child small CPL be road 3P.PL.POSS and people big
 tambié
 also
 ‘all children are on their [own] journey, as are the adults’

3 The genesis of generic-noun-based indefinites throughout Creole languages

3.1 Universalist explanations

The role of linguistic universals is traditionally acknowledged, by different approaches, as an essential part in the process of creolization. It is unclear, however, how exactly the role of universals in creolization is intertwined with that of superstratal and substratal influence and whether we are dealing with universals of first-language (L1) and/or second-language (L2) acquisition. The most famous approach that relates creolization to L1 acquisition is made by Bickerton: for this author, prototypical Creoles – those being in the center of the *Continuum of Creoles*¹² (Bickerton 2016 [1981]: vii–viii) – have formed as a result of the nativization of pidgins. The idea that pidgins somehow carry the germ of Creoles is an old one,¹³ yet Bickerton’s new approach consisted in making the process of creolization depend on the existence of a *bioprogram* – with which human beings are

However, there are three facts that make the latter interpretation rather unlikely. Firstly, there is no actual reason why the 2P object should be doubled (being expressed once in Spanish and once in PAL), especially in a (cleft) sentence that already has a focalized element (the subject *yo*) and even seems to introduce a secondary focus at the end (precisely, by means of *di belá belá*); in other words, there is no room for another sentence focus – the doubling of the 2P pronoun being the formal consequence of such additional focalization – i.e. a reading such as ‘I am the one who is telling you – and only you – the actual truth’ might just be too much. Secondly, despite the fact that preverbal object clitics from Spanish are often intertwined in PAL sentences (as much in the old recordings by Schwegler as in Lipski’s newer recordings: Lipski 2020: 86–87, 115–116), in such cases we would expect a Spanish conjugated verb (and not a clearly Creole verb, as in the example above). Thirdly, when the Spanish object clitic *te* is inserted in a PAL sentence, it usually corresponds to PAL 2P.SG pronoun *bo*, not to the pronoun *té* (< Sp. *usted*) (in the older recordings by Schwegler I find examples like *bo <te va[s]> agüé?* ‘are you leaving today?’, but not examples like *uté <te va[s]> agüé?*; the same is true for the examples presented by Lipski 2020: 86).

¹² Obviously, this concept is different from that of (post-)Creole continuum (DeCamp 1971).

¹³ In fact, Schuchardt (1888: 215) already defined *Jargon* – close to what we call today a pidgin – as “das Kreolische im Keim”, i.e. ‘a germinal Creole’.

genetically provided – whereby children who receive a pidgin as an input (L1) expand it naturally to a Creole. Bickerton developed the hypothesis of the bio-program in his book *Roots of Language* (2016 [1981]) and contextualized it later in a wider context in *Language and Species* (1990), in which he formulates a general theory of the human language. This theory is based on the distinction between *protolanguages* – a category in which he includes, alongside pidgins, the sign language of chimpanzees, the “talk” of human babies under two years of age, the talk of “Tarzans”, who did not receive any language input up to their adulthood, etc. – and *languages*, namely the natural, completely developed languages, including Creoles. Both books together analyze the syntactic properties that distinguish languages from protolanguages, properties such as recursivity, interpretation of null categories according to syntactic rules, etc.

For other authors, like McWhorter (2011), the language contact between adults results in a structural simplification, which brings about new restructured varieties that can develop into Creoles. Since they are younger languages, Creoles are structurally simpler than non-Creoles, in which complexity has been developing throughout history – often in connection with multi-secular written language cultures. It cannot be denied (no paradigm does that) that Creole languages are native languages (L1), but a great deal of the prototypically creole features are already present in the L2 or *learner varieties* of the adults, which are later passed on to subsequent generations. By stating, for instance, that “situations involving second language acquisition include classroom learning, language shift leading to the formation of “indigenized” varieties of [...] languages, and creole formation” (Winford 2008: 127), it is not implied that creolization is L2 acquisition, but rather that the latter process makes part of the former and, therefore, it makes sense to compare Creoles with other outcomes of L2 acquisition. In this regard, the discussion usually revolves around the role that children and their allegedly “creative” varieties play in the emerging creolophone communities – which is critical for Bickerton, but not for many other authors. This is the key to understanding how much (or how little) of the interlanguage features can be transmitted to the following generation of Creole speakers.

As for the most common developments of interlanguages, special attention has been paid to the overgeneralization of variants (generally, the analytic, more transparent ones) and the regularization of morphological irregularities (McWhorter 2011, Selinker 1972). As a matter of fact, such developments can account – partially, at least – for the phenomenon studied here: generic nouns are marginal but possible variants for conveying indefinite expressions in Ibero-Romance (see §3.2) and might therefore become general in the interlanguages, in which they prevailed over special indefinites. This change would give rise to a

bigger regularity of the system: “irregular” indefinite pronouns would be avoided in favor of the more regular indefinite NPs, in which common indefinite determiners are used. In addition, irregular morphology (such as the endings *-ie* and *-ien* of the Spanish indefinite pronouns *alguien* and *nadie nadie*¹⁴) is avoided in interlanguages too.

In reality, the formation of generic-noun-based indefinites can be explained, to a great extent, both by principles of L1 and L2 acquisition. Certainly, both processes were present in the primitive creolophone communities and, consequently, both played a role in shaping Creoles’ grammar. In fact, even from a Bickertonian perspective, some structures of the interlanguages can be inherited during the nativization and remain unchanged afterwards. I will briefly elaborate on this last issue.

As Bickerton himself acknowledges in *Language and Species* (1990: 164–196), some elements of *protolanguages* (like pidgins) can survive in *languages* (like Creoles), despite the fact that the change from one to the other is supposed to be mainly an abrupt one. For example, the author considers that the interrogatives formed on the basis of generic nouns (‘what-man’ for ‘who’, ‘what-thing’ for ‘what’, ‘what-place’ for ‘where’, etc.) in many Creoles are relics of their “pidgin phase” (1990: 183). In a protolanguage, it would have been practical to use one single non-referential element together with different (generic) nouns to form interrogative expressions.¹⁵ Furthermore, in the case of indefinites, we could formulate an analogous reasoning to justify the appearance of the analytic forms with THING and PERSON. In addition, such linguistic change would be taking place in a general context, in which indefinites are supposed to represent absolutely necessary information in protolinguistic communities (1990: 184–185) (protolanguages cannot simply leave the semantic category of indefiniteness open to contextual interpretation).

Moreover, the interpretation of null elements in protolanguages is not systematic (they rather require “guesswork identification” for their correct interpretation; Bickerton 1990: 169), which could have naturally triggered the overtiness of the nominal head, initially as a form of avoiding ambiguities with regard to the reference of the zero element: things/people that are mentioned in the previous discourse or can be identified situationally vs. things/people in general. Actually, if we assume the derivational link between pronouns and determiners

¹⁴The variant *nadien* is documented in Caribbean texts written by *semi-illiterate* Hispanic Creoles (white-descendants) during the colonial period (Gutiérrez Maté 2018: 546, 548, 589) (remember that Colonial Caribbean Spanish is the actual lexifier of PAL).

¹⁵Yet Bickerton is also aware that, on the other hand, generic nouns could have been too abstract for protolanguages (1990: 182).

(a framework that has a long tradition and has been applied even to account for the relationship between articles and personal pronouns: see Bosque 1989: 48–51, 179–191 and references therein), we can easily understand that this link extends to the case of indefinite expressions. As for the overtness of the phrase head, we have to acknowledge the fact that the syntactic configuration of the NP/DetP ([_{DP} [Det [_{NP} N (Adj)]]]) – as well as the configuration of syntactic phrases in general – is considered to be an elementary aspect of the transition from protolanguages to languages (Bickerton 1990: 191); thus, the explicit marking of phrasal structure (avoiding null heads) could also be seen as a natural reflexion of the bioprogram, especially when the overt morphological features of the determiner had already been lost during the pidgin phase (in PAL there is *un*, *aggú(n)*, *mucho*, etc. but not **una/unos/unas*, **alguna/algunos/algunas*, **mucha/muchos/muchas*, etc.) (possibly, the same principle accounts for the frequent use of THING and PEOPLE with non-pronominal NPs; see my comment on examples (18–19)). To illustrate this, we can notice the differences – and, most importantly, the structural correspondences – between the following indefinite expressions in Spanish and PAL:

Table 2: Phrasal structure of (pro)nominal indefinite expressions in Spanish and PAL

	nominal indefinite expression	pronominal indefinite expression
Example (constructed by linguist)	Sp: <i>He bebido mucha leche</i> PAL: <i>I a bebé mucho leche</i> Eng: ‘I drank lot of milk’	Sp: <i>He bebido mucho</i> PAL: <i>I a bebé mucho kusa</i> Eng: ‘I drank a lot’
DetP structure in Spanish	<i>mucha leche</i> : [DetP mucha [NP leche]]	<i>mucho</i> : [DetP mucho [NP ∅]]
DetP structure in PAL	<i>mucho leche</i> : [DetP mucho [NP leche]]	<i>mucho kusa</i> [DetP mucho [NP kusa]]

Of course, the same structural analysis can be extrapolated to other indefinite expressions analyzed throughout this chapter (like *to kusa* ‘everything’). The same DetP structure is, therefore, valid to account for both indefinite NPs and pronouns in both the Creole and its lexifier. Specifically, assuming that the generic noun in PAL occupies the same syntactic position as the null element (∅) in Spanish is key to understand the hypothesis.

In fact, we can even find languages/varieties emerged in (current or past) language contact settings that have developed the tendency towards the overt use

of nominal heads in pronominal expressions beyond indefinites (for example, in demonstratives). This is the case, for example, of the *quilombo* community of Jurussaca (Pará, Brazil), in which the proform *um/uma* is regularly used in demonstrative pronominals: *esse um/essa uma, aquele um/aquela uma* ‘that one’ (*esse/essa, aquele/aquela* in Standard Portuguese) (Campos & de Carvalho do Vale 2018) (this use resembles that of one in English and similar pro-forms in other languages: see also Haspelmath 1997: 29, 183–184 about ‘one’-based indefinite expressions).

In PAL, however, overt nominal heads (generic nouns) are mostly restricted to indefinite and, to a lesser extent, interrogative pronouns, i.e. we still regularly find $\emptyset$ in demonstrative, possessive and relative pronouns,¹⁶ which is possibly the main reason why language acquisition universals alone cannot account for the origin of PAL generic-noun-based indefinites. Most certainly, the type of changes studied in this chapter would not have taken place if the corresponding Spanish generic nouns had not occasionally been used with an unspecific (quasi-pronominal) reading, nor – most importantly – if Spanish had not come into contact with a language (Kikongo) that regularly employs generic-noun-based indefinites. The following two sections are devoted to these issues.

3.2 Superstratist explanations

The continuity between Creoles and their lexifiers has been highlighted and explained on a theoretical level predominantly – although not exclusively – by authors often referred to as “anti-exceptionalists” (cf. Mufwene 2001). In the particular case of Gallo-Romance linguistics, Chaudenson (whose impact on Mufwene is

¹⁶See the following examples:

- (i) (Fieldwork M. Gutiérrez Maté 2017)
 ese $\emptyset$ nu má jue-ba
 that $\emptyset$ NEG more be-IMP
 ‘it was only that’
- (ii) (Recordings by A. Schwegler 1985–1988)
 yo sí ten mailo nu pokke $\emptyset$ ri mí a morí
 1P.SG AFF have husband NEG because $\emptyset$ of me PRF die
 ‘I really have no husband, because mine has died’
- (iii) (Fieldwork M. Gutiérrez Maté 2017)
 kabeo liso, <dise> ma $\emptyset$ loke konosé-lo
 hair straight <say> PL $\emptyset$ REL know-3P
 ‘[she had] straight hair, say those who knew her’

evident) was the author who, together with his disciples, worked the most along this line of investigation (cf. Chaudenson 1992, 2003). From this perspective, Creoles derive from their lexifier languages, just as Romance languages derive from Latin. Specifically, Creoles are considered to have resulted from *approximations des approximations*. These took place, for instance, when the African slaves from the French plantations in the Caribbean did not learn French from the white colonizers (the plantation owners) but from the foremen, who spoke L2 French themselves.

According to the above, it is necessary to wonder about indefinites in PAL's superstrate: this leads us to investigate the use of the nouns *cosa* 'thing' and *gente* 'people' in 17th century Northern Colombian Spanish, which we can consider as the authentic lexifier of Palenquero. To this effect, I rely on the corpus of colonial Caribbean documents that I gathered and transcribed for my PhD thesis and with which I have worked since then.¹⁷ From this corpus of documents, I have selected those written in the governorate of Cartagena de Indias.

As regards the use of *cosa*, it is common to find it together with an indefinite (mostly, postnominal) determiner in cases in which the entire nominal phrase can function as an indefinite pronoun (a reading that is especially clear when the reference of *cosa* cannot be interpreted as a material thing). When the sentence polarity is positive, the determiner is *algún/a* (22), which is also possible when the sentence polarity is negative and, therefore, the interpretation of the indefinite expression is also negative (23); nonetheless, the negative determiner *ningún/a* can also be used with other overt markers of negative sentence polarity like the negation adverb (double negation) (24):

- (22) (Cartagena de Indias 1694, p. 45r, ls. 6–8)
se determino [...] que si se le aberiguara cosa alguna en este particular le diera la punición deuida a la naturaleza del delito
'it was decided that, if something was found out in this regard, it would be punished according to the nature of the crime'
- (23) (Cartagena de Indias 1695, p. 238v, l. 22)
No los he molestado en cosa alguna
'I have not bothered them at all' (/...in any way)
- (24) (Cartagena de Indias 1672; p. 82r, 19–20)
Embarcaron con orden de que no dejase sacar cosa ninguna del navío
'They embarked with an order not to let anything be taken out of the ship'

¹⁷ See Gutiérrez Maté (2013: 431–442) for a description of the documents and their archivist references.

In addition, *cosa* can be used with no determiner (25), which does not preclude the possibility that it is modified by adjectives (26):

- (25) (Cartagena de Indias 1694, p. 33r, ls. 9–10)
No hay cosa que más se pueda temblar que unas hiervas
 ‘There is nothing to fear more than some herbs’
- (26) (Cartagena de Indias 1693d, p. 5v, ls. 5–6)
Se han puesto de mi parte sin otra causa de que ven mi limpieza y que no hago cosa injusta
 ‘They have taken my side for no other reason than seeing my cleanliness and that I don’t do anything unfair’

Even in the absence of a quantitative study, it is clear that all these uses were also possible in other Spanish varieties at that time and even today (if anything happened to stand out in these examples, it would be the placement of the determiner *algún/ningún* in the unusual postnominal position, but this particular word order seems to play a role for the pronoun-like use of generic nouns in Spanish: most especially, a prenominal determiner *alguna* would not be compatible with the negative sentence polarity in (23)). As for the use of *gente*, there does not seem to be substantial differences with other Spanish varieties either. Especially as we do not register its use meaning ‘someone’ (the special indefinites *alguien* or *alguno* are used instead). However, a few interesting phenomena regarding the use of *gente* have to be noted: firstly, some uses in negative contexts are certainly close to the meaning of ‘no one’ (i.e. *gente* can be used instead of *nadie/ninguno*, which were, in any case, the predominant forms in Colonial Caribbean Spanish):

- (27) (Cartagena de Indias 1772, p. 513v, ls. 3–5)
volvió a salir a dicha Casa la que encontró en silencio por haver reconocido no haver gente dentro de ella
 ‘He returned to enter the aforementioned house, which he found in silence as he realized there was no one inside.’

Secondly, when *gente* was used with a non-arbitrary meaning, its reference adapted to the cultural and sociological idiosyncrasies of the colonial Caribbean: for example, *gente* was commonly used for ‘militias’ against the enemies of the city and, in a more general fashion, to refer to certain social and/or ethnic groups, in which the speaker could include him/herself or not. Thirdly, in relation to what has just been said, *gente* could adopt the meaning of a generic subject and even a generic subject with inclusion of the speaker (Gutiérrez Maté 2013: 80–82),

which resembles the use of a first person plural pronoun. Nevertheless, despite the fact that acknowledging that speakers used *gente* to include themselves in the predication may well be revealing as regards the origin of some current uses of PAL *hende* (see examples 16–17), it does not say anything about its use as an indefinite pronoun. In other words, the linguistic change “generic noun > indefinite pronoun” is different from the change “generic noun > generic pronoun > 1P.PL pronoun”, even though one process does not prevent the other from taking place, as has actually happened in the case of PAL (not so in the case of Caribbean Spanish, since the marginal use of *gente* as (something like) a 1P.PL pronoun does not seem to have really taken root, nor did it develop further after the colonial era).

Finally, the noun *persona* deserves a separate comment, since it exhibits – more frequently than in the case of *gente* – semantic readings close to that of an indefinite pronoun, especially in negative contexts (the phenomenon, evidently, is well known in other world languages, including French):

- (28) (Cartagena de Indias 1694a, p. 3r, ls. 1–4)
Domingo Criollo, con notiçia que havía tenido de dichos bandos, havía publicado uno en su palenque San Miguel, para que no saliesse negro d[e] él, ni tubiesse comunicación con persona
‘Domingo Criollo, after having known about the aforementioned town proclamations, had announced one in his *palenque* called San Miguel, so that no black person would leave it or have any communication with anyone [outside]’

However, as we have seen, *persona* does not seem to have entered (Traditional) Palenquero, neither in its canonical use as a noun nor in its potential use as an indefinite pronominal.

As it turns out, the superstratist explanation alone cannot account for the emergence of PAL generic-noun-based indefinites either. Moreover, if the drift towards generic-noun-based indefinites were somewhat inherent in the Spanish language, we could not explain why these have not become the main strategy for building indefinite expressions in other Hispanic varieties, not even in today’s Northern Colombian Spanish.

3.3 Substratist explanations

3.3.1 The data from the APICS

Today it is easy to prove that generic-based-noun indefinites are predominant in Creole languages; they clearly outnumber interrogative-based indefinites, which

constitute the dominant group throughout the world languages. Table 3 compares the data of the WALS, which contains a sample of 326 languages, with the data of the APICS, which codifies 74 languages (most of them being Creoles, together with a few other varieties resulting from language contact: mixed languages and partially restructured varieties). The APICS data are, in fact, taken from the WALS-like APICS (<https://apics-online.info/wals/21#2/30.4/9.8>), in which some values have been adapted in order to make them compatible with those that had previously served to codify the interlinguistic variation in the WALS: on the one hand, a “mixed” value is introduced in those cases in which the APICS specified the actual percentages of such “mixtures” (linguistic variables); on the other hand, the values “generic-noun-based indefinites” and “old generic-noun-based indefinites continuing somebody/something”,¹⁸ which were distinguished in the APICS, are unified in the WALS-like APICS. The resulting data speak for themselves: whereas in non-Creole languages the proportion of generic-based-noun indefinites is approximately $\frac{1}{4}$, the proportion rises to about $\frac{3}{4}$ in Creoles. As is well known, the four basic values of the WALS (interrogative-based, generic-noun-based, special and existential constructions) rely on the distinction first introduced by Haspelmath (1997).

Initially, it may seem that Creoles naturally tend to form generic-noun-based indefinites, so that we could think of these as universals of creolization (in line with the arguments presented under §3.1). However, a closer look at the maps of the WALS and the WALS-like APICS shows that this is not the case. We can, for instance, observe three linguistic areas in which, according to the WALS, we find different strategies for the building of indefinite pronouns:

- 1) Sub-Saharan languages (mostly from the Niger-Congo family), which belong to the areas III and IV according to the classification of African languages by Güldemann (2010). In these languages (marked in blue in Figure 1), the use of generic nouns prevails.
- 2) Southern Indian languages (from the Dravidian family). In these languages (marked in red in Figure 1), the use of interrogative-based indefinites prevails.
- 3) Philippine languages (as a relatively homogeneous subgroup within the Austronesian languages). In these languages (marked in white in Figure 1), the use of (pseudo)existential constructions prevails (when this strategy is combined with other types of indefinites, the languages are marked in grey in Figure 1).

¹⁸This was the value for those cases in which generic-based-noun indefinites were already found in the lexifier, as in the case of English *something*, French *quelque chose*, etc.

Table 3: Indefinite pronouns in the world languages and in Creoles

	WALS (world languages)		APICS (world Creole languages and a few other language contact varieties)	
	N	%	N	%
Interrogative-based (<i>wh</i> -indefinite)	194	59.5%	4	5.4%
Generic-noun-based (THING, PERSON)	85	26%	54	72.9%
Special (different lexemes or different morphology)	22	6.7%	6	8.1%
Existential constructions ('there is one who...' meaning 'someone...')	2	0.6%	1	1.3%
Mixed (combination of two or more of the former)	23	7%	9	12.1%
(language sample)	326		74	

For our purposes, the most interesting observation is that these three groups of languages also constituted the substrates of different groups of Creoles, which – with the exception of Group 1a (s. Figure 2) – formed *grosso modo* in the same areas. If we compare the world languages from Figure 1 to the Creoles from Figure 2, there seems to be no doubt about the substratal influence in the making of indefinite pronouns in Creoles. As for the particular case of the group 1a (Caribbean Creoles), it is well known that they came into being with participation of the same substrate languages that influenced group 1b: Caribbean Creoles emerged as a consequence of forced migration (slave trade) of speakers of Niger-Congo languages, who made contact with other exogenous languages (those spoken by the European colonists) in the New World. In addition, the few yellow spots in the area 1a (representing Creoles with special indefinites) should be taken carefully or even partially recoded; for instance, indefinites in Papiamentu (in yellow in Figure 2) should be considered “mixed” rather than “special” for two reasons: (1) the forms *un hende* (Kouwenberg 2013: ex. 47–42) and *algun hende* (Maurer 1998:

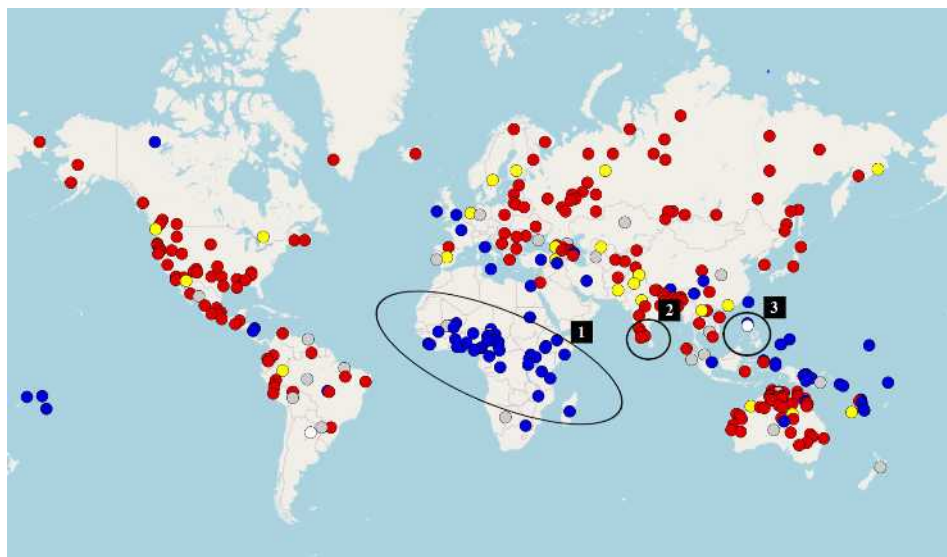


Figure 1: WALS map, feature 46A (“Indefinite pronouns”). Base map © OpenStreetMap contributors.

60) are consistently used as personal indefinite pronouns, and (2) the form *unkos* (Maurer 1998: 60) can also be used for the meaning of ‘something’ alongside the special indefinite *algo* (Kouwenberg 2013: ex. 47–52).¹⁹

The above does not mean that the formation of indefinite expressions in *all* Creoles is directly inherited from their respective substrates, since the emergence of each Creole is idiosyncratic according to various factors: number of the substrate languages, homogeneity/heterogeneity of the typological characteristics of the substrates, degree of exposure to the lexifier language, and many other aspects concerning each particular language ecology. It is, however, evident that the preferred structural type of indefinite pronouns in many Creoles coincides with that of their substrates, which obliges us to take the hypothesis of substratal influence into account. Having made these clarifications, the relationship between Creoles and their substrates as regards indefinite pronouns is depicted in Table 4.

The Creoles of the groups 2 and 3 have preferred the indefinite pronouns type of their substrates over the type displayed by their lexifiers. For example, the lexifiers of both Sri-Lankan Creoles do not use interrogative-based-indefinites:

¹⁹The origin of *algo* in Papiamentu could be attributed to 17th century Ibero-Romance or to relexification by means of the canonical (Caribbean) Spanish form *algo* at a later evolutionary stage of the Creole.

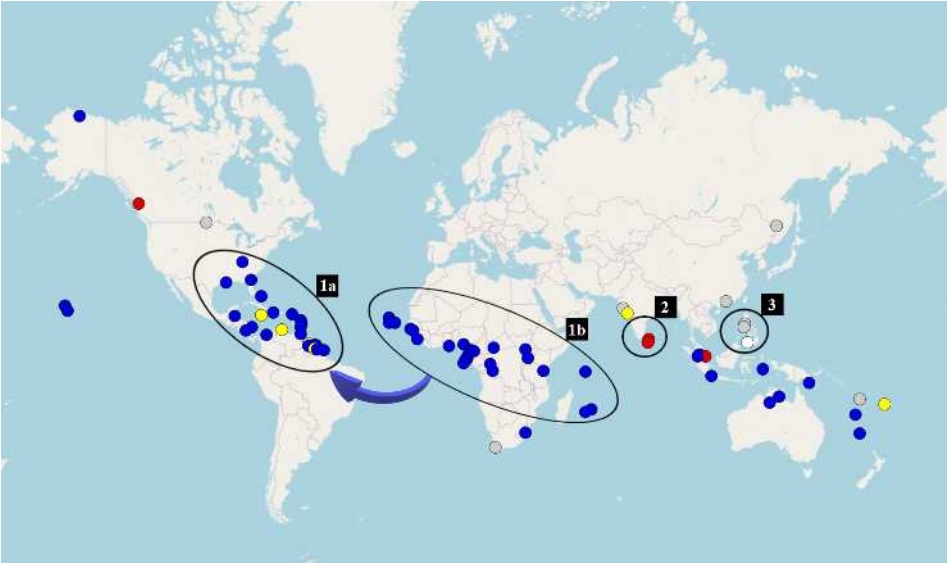


Figure 2: WALS-like APICS map, feature 21 (“Indefinite pronouns”).
Base map © OpenStreetMap contributors.

Table 4: Indefinite pronouns in the world languages and in Creoles

Substrate languages (linguistic areas)		Creoles
Group 1. Areas III and IV of African languages according to Güldemann (2010)	⇒	Group 1a. Caribbean Creoles Group 1b. African Creoles (both continental and insular)
Group 2. Southern Indian (Dravidian) languages	⇒	Group 2. Sri Lankan Creoles (Sri Lankan Portuguese, Sri Lankan Malay)
Group 3. Philippine languages	⇒	Group 3. Chabacano varieties (most especially, the variety of Zamboanga)

Malay makes use of generic nouns, whereas Portuguese combines special indefinites (*alguém*) with generic-noun-based (*alguma coisa*). As regards Chabacano, it becomes evident that the use of existential sentences with the meaning of indefinite expressions is not inherited from the lexifier (Spanish); in addition, this feature is, from a typological perspective, extremely unusual, so the likelihood that Chabacano (a Philippine/Spanish Creole) had developed this very feature “on its own” – relying on universal principles – are also extremely low.

Many Creoles spoken in the area 1b do not allow for the impact of the substrate languages to be proven, since both substrates and superstrates coincide in using generic-noun-based indefinites (for instance, Lingala has Bobangi, a Bantu C language, as its lexifier and other Central and West African Languages as its substrates – cf. Meeuwis 2013, where both Bantu languages and Niger-Congo languages spoken in West Africa generally employ generic-noun-based-indefinites). As regards those Creoles spoken in the area 1a, we find a somewhat more relevant impact of the superstrates (remember the case of Papiamentu *algo*), but substratal influence should be considered to be the most relevant factor in the formation of indefinite expressions throughout Caribbean Creoles. As depicted in Table 3, generic-noun-based indefinites are not strange in universal terms ($\frac{1}{4}$ of the languages of the world employ it), but the extremely high percentage of its appearance in Creoles (up to $\frac{3}{4}$ of these languages) does not only seem to be the effect of universal dynamics, yet either the substrate or the superstrate must have played the decisive role. In the case of PAL, where the superstrate makes a rather marginal use of generic nouns as indefinites, substratal influence should be considered the ultimate trigger of the linguistic change analyzed here.

3.3.2 The Kikongo data

Unfortunately, Kikongo is not represented in Figure 1. However, this language follows the general tendency found throughout Sub-Saharan languages of using generic nouns to form indefinite expressions. In the words of Kyala (2013: 118): “Na ausência de palavras precisas para exprimir a noção exacta de pronome indefinido, o kikongo faz recurso às locuções pronominais indefinidas” (‘Due to the lack of precise words to express the exact notion of indefinite pronoun [= special indefinites], Kikongo makes use of indefinite pronominal expressions [= generic-noun-based indefinites]’). In Kikongo, there are in fact various words with the meaning of THING that can adopt an indefinite interpretation: The most common are *kyuma* ~ *kima*, *diambu*, and *kimvela*, although there are others as well (cf. Laman 1912: 150–151). *Kyuma* seems to most frequently have the concrete

meaning of ‘object’, which means that, when it is used in indefinite expressions, its potential referents are also material:

- (29) (Fieldwork M. Gutiérrez Maté 2020)

mwisi kyuma ko

mu-isi ki-uma ko

NC3-inhabitant.from NC7-thing NEG

‘there is nothing inside’ (lit. ‘nothing is from here’) (answering the question: *what do you have in the box?*)

Diambu, when used as generic noun, can have a more abstract value than *kyuma* and be often translated with ‘problem’. Because of this, the potential referents of *diambu* can also be more abstract (example (30) is the answer of an informant of Kiyombe when I asked him to translate *eu vou te dizer algo* ‘I’m going to tell you something’ from Portuguese to Kiyombe):

- (30) (Fieldwork M. Gutiérrez Maté 2020)

minu diambu iakukamba

minu di-ambu i-a-ku-kamb-a²⁰

1P.SG NC5-thing/problem 1P.SG-PRES/FUT-2P.OBJ-tell-FV

‘I am going to tell you something’

Lastly, *kimvela* is the word that has the most ambiguous status of the three, perhaps being mostly restricted to its use as a pronoun (with nominal morphology, where *ki-* is a prefix of class 7, the same as we find in *kyuma*). It is defined by Laman (1936: s.v.) as “pas un seul; pas un brin; rien; néant”, i.e. exclusively as a negative indefinite (‘nothing’). However, at least Kiyombe speakers (keep in mind that Laman describes a different dialect, which belongs to the Central Kikongo sub-clade: Bostoen & de Schryver 2015: 147) also use it as a positive indefinite (‘something’). As a matter of fact, when my informants were asked to translate this word into Portuguese, they answered primarily with the positive indefinite (*alguma coisa* or *algo*):

- (31) (Fieldwork M. Gutiérrez Maté 2020)

twala kimvela

twala ki-mvela

bring(IMP) NC7-thing

‘bring something!’

²⁰Future tense is not a common meaning of the circumfix *a-...-a* throughout Kikongo dialects, but it can be found in some varieties and has even been described previously by missionary grammars (cf. Dom & Bostoen 2015: 170) and references therein.

As for personal indefinite pronouns, Kikongo uses the generic noun *mu(n)tu* ‘person’ (PL *ba(n)tu*):

(32) (Kyala 2013: 117)

Etata, muntu mosi ka wiza kunzo
 e-tata mu-ntu mosi k-a-(k)wiza ku-nzo
 AUM-father NC1-person one 3P.SG-PAST-come NC17(LOC)-house/home
 ‘Father, someone had come home’
 (Port. translation by the author: *Pai, tinha vindo alguém (alguma pessoa) em casa*))

(33) (Carter 1999: 40)

kavàkala muntu ko
 ka-va-a-kala mu-ntu ko
 NEG-LOC(NC16)-PAST-COP NC1-person NEG
 ‘there was no one’

When used as a positive indefinite (‘someone’), *muntu* can be modified by the determiner *mosi* (‘one, same’): *muntu mosi* has even been described as the canonical form for ‘someone’ in some grammars (Kyala 2013: 52–54), but *muntu* can also appear without further modification (according to some examples by Laman 1936: s.v. *muntu*). In the French-Kikongo dictionary by Biyoko Mabua (2017) *mutu mosi* is consistently translated as *une personne* (‘one/some person’), whereas *mutu* (without determiner) is translated as the actual indefinite pronoun (*quelqu’un*). When *muntu* is used as a negative indefinite (‘no one’) the generic noun seems, at first, to generally lack any determiner (this was so, at least, amongst my informants), but there is at least one source that indicates the two variants: *ka muntu ko* and *ka muntu mosi ko* (see Dereau 1957: s.v. *personne*). In a similar way to PAL (s. examples 20–21), attributive adjectives in copular sentences and similar constructions seem to often be accompanied by an explicit nominal head (a generic noun):

(34) (Biyoko Mabua 2017: 150)

Bibila kima kimboti
 Bibila ki-ma ki-mboti
 Bible NC7-thing NC7-good
 ‘The Bible is good’ (/‘The Bible is a good thing’)

It is important to note that, just as today’s Kikongo, 17th century Kikongo (the actual substrate of PAL) seems to only make use of generic-noun-based indefinites. In this paper, I cannot elaborate on these problems, which would demand

a detailed analysis of the different Kikongo *doculects* written in the 17th and 18th centuries (Bostoen & de Schryver 2015, 2018), but a look at the very first source written in Kikongo (if we leave aside the few quoted words or sentences in this language that can be found in official documents written in Portuguese during the 16th century) can quickly confirm the structural tendency outlined above. I am referring to the *Doutrina Cristã*, a bilingual catechism Portuguese → Kikongo, published in 1624: even though there are no cases of *alguém*, *algo* or *nada* in this text, we do find several instances of *ninguém*, which are regularly translated as *muntu* ‘person’ (in negative sentence contexts) (s. the edition by Bontinck 1978: 119, 145, 195).

Finally, it has to be noted that the grammaticalization of *muntu* – or its cognates in other closely related Bantu languages – as a sort of focus particle²¹ does not seem to affect the vernacular varieties of Kikongo that I am dealing with (if anything, the change would find itself at its very early stages in Kiyombe). Such development has taken place in some languages of the Bantu B, C and H areas (van der Wal & Maniacky 2015), but is especially characteristic of two vehiculars, Kituba and Lingala, which have often been classified as Creoles. Since these languages probably came into being as late as the second half of the 19th century or at the turn of the 20th century, we can easily understand that the linguistic change “generic noun PERSON > focus particle” has not affected PAL in any way.

4 A first account of the use of indefinite pronouns in Cabindan Portuguese

Angolan Portuguese consists of a set of very heterogeneous varieties. Consequently, its analysis can be approached from many points of view (including the question of pluricentrism, i.e., by wondering about the possible endocentricity of Angolan Portuguese within the Lusophone World). In my case, however, it is only the “fossilized”²² L2 varieties of Portuguese spoken by bilingual speakers with Kikongo (especially, Kiyombe) as a predominant L1 that are taken into account. All the examples I will present in this section were uttered by bilingual speakers.²³

²¹The change included several evolutionary stages of type “John is the PERSON who did it” > “John PERSON did it” > “The cat PERSON did it”, etc.

²²See Roche (2013), Selinker (1972: 82–86).

²³Many examples come from old informants from the Cabinda province who learned Portuguese during the colonial period in a non-monitored way (although there was certainly contact with the European standard variety in their short time at school); thus, to give an example, one

Especially after the colonial period, Portuguese became the authentic lingua franca in the country for several reasons.²⁴ It started being regularly used by almost the entire population in the cities and by a constantly increasing number of people in the countryside. Such *vehicularization* of Portuguese did not necessarily bring the generalization of a given variety, nor the formation of a new one through koineization (at least not yet), even though there certainly are some features that have been widespread all over the country and are therefore characteristics of what we can today call “Angolan Portuguese”. Amongst these features, just to mention one that falls within the scope of indefinite expressions, we find the use of *bué*, an indefinite determiner (and elative adverb) of Bantu origin,²⁵ which has even spread out to other lusophone areas, including Portugal, as a marker of “youngspeak” (Almeida 2008):

- (35) *Eu gosto bué/Lá tem bué de candongueiros*
 ‘I like it very much’/‘There are many candongueiros [‘informal shared taxis’] there’

of the communities where I have done extensive fieldwork (Lites, municipality of Buco Zau, Cabinda) consisted of an old colonial *fazenda* where the employees had learned Portuguese due to their contact with the white Portuguese settlers who owned the farms. Under these circumstances, Portuguese-based interlanguages, characterized by morphological simplification, the overgeneralization of analytic variants and various transfers from Kikongo, fossilized. It is in this sense that we can speak of the *partially restructured* varieties of Portuguese that are spoken in Cabinda or, in a more general fashion, in Angola (cf. also Inverno 2009, who also uses this concept with reference to the Portuguese varieties from Dundo, in the province of Lunda Norte).

²⁴One of the factors that had great impact on the generalization of Portuguese was the Angolan civil war (1975–2002). During this conflict, thousands of people had to migrate from one province to another (where a different indigenous language was spoken), which made the use of Portuguese more convenient.

²⁵The Umbundu etymology has been proposed on several occasions (cf. Schmidt-Radefeldt 2010), but other source languages are also possible in theory, including Kikongo. The problem cannot be easily solved, since the possible sources of *bué* that we are dealing with are formally close cognates. For instance, in Kiyombe, the indefinite determiner meaning ‘many’ is *-phwedi*, as in the example below. A linguistic change Kiy. /phwedi/ > Port. /bwe/ seems also entirely possible:

- (i) *tsinzau tsiphwedi tsidi mu Afrika*
 ziN-zau zi-phwedi zi-idi mu Afrika
 NC10-elephant NC10-many NC10-COP LOC(NC18) Africa
 ‘there are many elephants in Africa’
 (fieldwork MGM, exercises of translation Portuguese → Kikongo)

For the most part, however, the perception of Angolan speakers clearly distinguishes between different *sotaques* ‘accents’, which, in reality, goes way beyond prosodic regional differences. Speakers can easily identify the (Umbundu-influenced) *sulano* variety, the (Kikongo-influenced) variety of the Bakongo people from Northern Angola, the specific *calão* spoken in Luanda (a folk concept under which two different realities can be understood: the Kimbundu-influenced varieties spoken in the city (cf. Mingas 2000) and the varieties developed in the *bairros sem identidade*, i.e., suburbs with no clear ethnic background like Lixeira-Sambizanga and others), etc. Accordingly, the linguistic features resulting from partial restructuring can also be different from one region to another, even though some commonalities are expected, insofar as the typological characteristics of the different *línguas nacionais* spoken in Angola (mostly belonging to the Bantu H, K and R groups) are, to a great extent, the same.

Creole languages never seem to have emerged in Angola (see Gutiérrez Maté 2020: 112–117 and references therein). Therefore, we expect differences between the two Ibero-Romance/Bantu vernaculars contrasted here, i.e. qualitative and quantitative differences that serve to determine the “degree of restructuring”. On the one hand, some features that were transferred from Kikongo to PAL do not make it into CP; on the other hand, some features that are transferred to both PAL and CP became grammar rules in the former, whereas they are just tendencies of use in the latter. The use of generic-noun-based indefinites can illustrate the latter principle, since it is by no means as systematic in CP as they are in PAL. Lastly, in Angolan Portuguese (including CP), we can find other phenomena that are primarily related to borrowing: most especially, amongst those speakers who employ Portuguese more frequently, the Kikongo influence is not so much the result of transfer (from the L1 to the L2) as the result of borrowing (from the L2 to the L1) (see Thomason & Kaufman 1988 about both directions of change in language contact scenarios). According to this idea, we can find some lexical and grammatical loanwords from Bantu languages, including the above-mentioned use of *bué*, in CP as well as in other varieties of Angolan Portuguese. This phenomenon has to be distinguished from the one I am analyzing in this work, even though there may be some overlapping areas between the two phenomena: for example, if generic-based nouns were proven to be used more frequently in Angolan Portuguese than in other varieties of Portuguese, we would still have to decide whether this use results from transfer or from (structural) borrowing. In most cases, it is only the contact ecology and the particular type of bilingualism that allow us to solve the problem: in the particular case of my informants, who are more used to speaking Kikongo at an in-group level, one would assume that a higher frequency of generic-noun-based indefinites in Portuguese would be a consequence of transfer from Kikongo.

The variation between *coisa* (the generic-noun-based indefinite) and *algo* (the special indefinite) can be observed in an example like the following, recorded in a “paragem sem nome” (‘nameless stopover’) a few kilometers from Lândana (Cabinda):

- (36) A (to B): *essa menina vai vir a que hora?*
 ‘At what time will this girl come?’
 B (to C): [...] *lhe mandaste...?*
 ‘you sent her...?’
 C (to B): *foi comprá aí um coisa*
 ‘she went there to buy something’
 B (to A): *foi comprar algo aqui*
 ‘she went there to buy something’

In this case, both C (speaking to B) and B (speaking to A) refer to the same fact: the girl that A is asking for went to a nearby shop to buy something, which, deliberately, is left unspecified (it can be inferred from the context that it is a purchase of groceries that she repeats with a certain regularity). Thus, there is the same referent, yet an informant uses *um coisa* (in which case the use of the masculine indefinite article instead of the feminine one represents another typical restructuring phenomenon), whereas the other informant employs *algo*. It is probably no coincidence that the use of *um coisa* takes place in the (in-group) communication between C and B (husband and wife respectively, both middle-aged), whereas *algo* appears in the (out-group) communication with A, the foreign interviewer.

Furthermore, the example is interesting because it demonstrates that the use of generic *coisa* can be introduced by the indefinite article (*um[a]* ‘a’) and not only by the indefinite determiner (*alguma* ‘some’), which is anyway possible in CP as an alternative to *algo*. As a matter of fact, *algo* “is considered archaic and pragmatically highly marked in (Modern) Standard Portuguese” (Cardoso 2013: ft. 21), so it has mostly been substituted by the generic-noun-based expression *alguma coisa* (and also *qualquer coisa*) ‘something’. As we have seen, this is also the reason why, as for the preferred strategy for building indefinite pronouns, the Portuguese language is classified – accurately so – as “mixed” in the WALS (see Figure 1). In Cabinda, *algo* seems to be perceived as a more correct and less vernacular option than (*alg*)*uma coisa*.²⁶ Possibly, it is perceived as a more polite

²⁶Whether the use of *algo* is the result of an “archaism” or an *idiomatization* (cf. Koch 1997) out of official discourses – or both – still remains to be clarified: as stated before, Angolan Portuguese is many things, and it would not be surprising if some uses that are restricted to a few discourse traditions in Portugal had become more accepted in Angola through administrative documents, political discourses and the media – which are often oriented to *written* European Portuguese.

form too (the following example was produced by a youngster who was trying to apologize on behalf of his grandfather when the latter was declining to be interviewed in a rather impolite manner):

- (37) *desculpa o se[nh]o[r], queria so falar algo, é que [...]*
'Excuse me, sir. I just wanted to say something. The thing is [...]

Further research should be able to determine the semantic nuances of all the possible variants that seem to convey the meaning of 'something' in CP, as well as the structural, stylistic and sociolinguistic factors that account for the use of one or the other variant. Alongside those forms that have already been mentioned (*um coisa, uma coisa, alguma coisa* – possibly *algum coisa* too – and *algo*), we also find the generic noun with no article/determiner (*coisa*) and the hybrid form *um algo*. For example, the following sentences (quite similar in content) were recorded in an interview with the same speaker (male, 38, Buco Zau, Cabinda):

- (38) *você tem coisa para falar?*
'do you have something to say?'
(39) *tem um algo a dizer*
'there is something to say'

However, in this chapter, we focus on the use of those variants that are structurally closer to the grammar of the substrate, like the first example of (36) and, most especially, (38), in which *coisa* is not modified by any indefinite article/determiner.

Just as we saw in the case of PAL, the generic noun with the meaning **THING** can also be used to form other indefinite pronouns in CP. Even though I have never heard *toda coisa* (instead of *tudo*), the indefinites *pouca coisa* (40) and, most especially, *muita coisa* (41–43) are very frequent:

- (40) *já tem setenta e dois ano que vou fazer, não é pouca coisa*
'I am already about to be 72 years old, that's no small thing'
(41) *naquela altura [...] eu trabalhei muita coisa, fiz isso, fiz aquilo...*
'at that time I worked a lot, I did this, I did that...'
(42) *quem fez a quarta classe [...] quer dizer que entende muita coisa*
'those who completed the fourth class [...] it means that they were able to understand a lot [of the Portuguese language]'
(43) *já é muita coisa!*
'that's too much!'

Interestingly, we often find *muita* instead of *muito ~ muita coisa* :

- (44) *dialeto é muita, agora o português so um bocado*
 ‘the dialect is a lot [=Kiyombe I know a lot], but I only know a little Portuguese’
- (45) *papá não sabe muita, não sabe não*
 ‘dad doesn’t remember much [about that time], he just doesn’t remember much’

Although it cannot be ruled out that the form *muita* results from the neutralization of gender morphology, this hypothesis is rather unlikely, since my data clearly suggest that the masculine form prevails over the feminine in those cases in which the marking of grammatical gender has been restructured (as in the case of *um coisa*). An alternative explanation is that *muita* results from the routinization (grammaticalization) of *muita coisa* in the restructured varieties of those speakers whose bilingualism leans towards regional Portuguese. I get this idea, partly, from examples like (45), a real case of an interview with an old father (91 years) and his son (60 years) in the ethnic neighborhood of the Bassolongo in the city of Cabinda (the Bassolongo are the indigenous group from the province of Zaïre, in this case from the city of Soyo): Whereas the father, who learned Portuguese during the colonial period, alternated between generic-noun-based indefinites and special indefinites, the son also produced examples like (45). Although both father and son are bilingual, it can be assumed that the latter was always a much more active “user” of Portuguese, since he did not really get to use Kikongo in his everyday Cabindan life (his family dialect, Kissolongo, is different from the one that is predominant in the city of Cabinda, Iwoyo). This hypothesis – which I note in a provisional way here – would consist in the following chain of changes: (1) *muito* > (2) *muita coisa* > (3) *muita* , where (1) would be the Standard Portuguese form, as learned at school today and at colonial times, (2) would be the prototypical variant resulting from language restructuring under the influence of Kikongo and (3) the variant used in the actual nativized variety of CP, which has – partially, at least – come into being out of the Kikongo-influenced restructured varieties of Portuguese.

Unlike PAL *kusa* , CP *coisa* does not seem to be used with a negative meaning (which does not mean that this use will be registered one day). What can in fact be often found – at least, among older informants – is the use of the NP *um(a) coisa* in emphatic negative contexts (related to counter-expectation):

- (46) *oh: não vou te dizer uma coisa*
 ‘well, I am not going to tell you anything’

- (47) *não trouxe uma coisa para mim?*
'did you not bring me anything?'

In these examples, the speaker is not only negating the propositional content of the sentence but also an inference: by using *uma coisa* in (46), the speaker is explicitly contradicting the assumption made by the interlocutor (myself) that he was willing to tell me some anecdote of his life. As for the speaker of (47), he is not just asking me whether I did bring something or not, but also emphasizing that, under the circumstances of the conversation (and being already halfway through the conversation), I should in fact give something (a gift, or money).

For the most part, however, the indefinite with the meaning of 'nothing' in CP is *nada*, i.e. the special indefinite that constitutes the canonical negative indefinite expression throughout Portuguese varieties:

- (48) *ta rir é por qué? você que não sabe nada*
'why are you laughing? You are the one who knows nothing'

It has to be noted, though, that *nada* can also be used in ways that are far from Standard Portuguese: these include its use as an "anaphoric" extra-sentential negator (= 'no') (Bosque 1989: 51–52, Zanuttini 1990), as in answers to either positive or negative questions (s. examples (49) and (50), respectively), and its use as a second sentential negator (= 'not'), which is placed in the sentence-final position (examples 51–52). The latter may well be derived from the former and builds a special type of negation pattern, which is not only an alternative to the canonical Ibero-Romance negation pattern (preverbal negation), but also to the double negation of the type "*não* + V (+ X) + *não*" (this is also found in some Cabindan speakers and has been extensively described in other Ibero-Romance varieties like, most notably, Brazilian Portuguese and Palenquero; Schwegler 2016b, Schwenter 2016). Utterances with *nada* appear to be somewhat more emphatic than those with *não*. However, a closer look at the functional limits between both negators, *não* and *nada*, as well as the study of the reasons that led to the emergence of this very dichotomy of uses would be far beyond the scope of this chapter.²⁷

²⁷Suffice it to say that I consider this linguistic change to be related to language contact (in our particular case, with Kikongo, which formally distinguishes between anaphoric 'no' and sentential 'not' and also has double sentential negation) as well as to natural outcomes of L1-acquisition (after all, in the first stages of L1-acquisition there is only extra-sentential negators, which may develop later into sentential negators; Cameron-Faulkner et al. 2007).

- (49) A: *uma antiga doença era o beri-beri, ouviram alguma vez beri-beri?*; B: *nada!*
 ‘A: *beri-beri* was an ancient disease, have you ever heard *beri-beri*?; B: no!’
- (50) A: *Aqui não havia portugueses?* B: *na:da! havia, mas quer dizer eles aqui não pagava[m] imposto*
 ‘A: Were there no Portuguese here? B: no! there were some, but I mean that they didn’t have to pay taxes here’
- (51) *mas aquela pessoa lhe deram tiro, não é? não morreu nada!*
 ‘[in that telenovela] they shot that guy, right? but he did not die!’
- (52) A: *não seria “kibanga” em dialeto?*; B: *esta palavra eu não ouvi nada*
 ‘A: would that not be called *kibanga* in your dialect [=Kisundi]?; B: this word I have not heard’

I will not elaborate here on the multiple uses (some more grammaticalized than others) that *coisa* can adopt outside the domain of indefinites. However, I would like to note that *coiso*, which is well known in many Portuguese varieties for referring to objects or people that speakers cannot – or do not want to – name,²⁸ is extremely common in Angola (including Cabinda), where it has also developed as a sort of hesitation marker:²⁹

- (53) *município de... município de... tangente, de... coiso de... como lhe chamam aí?... de... não sei... de Cabinda... não é?... é me[s]mo de... de Cabinda*
 ‘municipality of... municipality of... bordering on... like... how is it called? not sure... Cabinda, right? it is just Cabinda’

Whether this particular use (which is certainly known in other varieties of Portuguese and Spanish (*coso*), but maybe not to the exact same extent as it is in Angola) can be somewhat related to Kikongo and/or other Bantu languages requires further research.

As regards personal indefinites, we again find variation between generic-noun-based and special indefinites (*alguém*) in CP. The generic noun that adopts the indefinite reading is *pessoa* ‘person’ and not *gente* ‘people’ (as has been shown in

²⁸The derivation of *coiso* and even *coisar* as a verb (‘to make something’), sometimes with sexual connotations, gives an idea of the wide spectrum of uses of the generic noun *coisa* in many Portuguese varieties.

²⁹Example (53) was registered in naturalistic speech. It is noteworthy, too, that the examples that my informants made up when asked about the use of *coiso* consisted, first and foremost, of its use as a hesitation marker, like *Eu estava falá com... coiso... coiso... o Miguel!* ‘I was speaking to...eh.. eh... Miguel!’

the cases of PAL and Kikongo, the limits between the indefinite and the generic reading are sometimes blurred; see also (56) further below). Even though the examples (54) and (55) are relatively similar as regards their propositional content and fit well into an existential quantificational reading (in both cases, it is about finding someone who can or cannot speak a given language), we also notice that, in some contexts at least, *pessoa* can alternate with *um(a) pessoa* quite freely (the data collected so far cannot help us to distinguish the precise semantic nuances of each form):

- (54) *é difícil você encontrar um pessoa da República Democrática do Congo conversar em francês*
 ‘it is difficult that you find someone (/a person) from the Democratic Republic of the Congo talking in French’
- (55) *de vez em quando vais encontrá pessoa que entenham [sic] português, mai... mai só falem fiôte*
 ‘sometimes you are going to find someone/people who understands Portuguese, but only speaks Fiote (=general designation for the Cabindan dialects of Kikongo)’

The preference of *pessoa* over *gente* for indefinite expressions is not surprising from a universalistic perspective, but it is in fact surprising inasmuch as *gente* is the selected generic noun for the making of indefinite pronouns meaning ‘someone’ in nearly all Niger-Congo/Portuguese Creoles (cf. Cape Verdean *algum djenti* < Port. *alguma gente*, Sãotomense *ũa ngê* < Port. *uma gente*, etc.), including Papiamentu (even though in this case the generic noun seems to have been relexified by its cognate *gente* /*hente*/ in (Caribbean) Spanish: Pap. *un hende* ‘someone/somebody’³⁰) (Haspelmath & The APiCS Consortium 2013). In CP, *gente* keeps being used as a kind of arbitrary generic noun, as in Standard European Portuguese. It is also interesting to note that, when my informants were asked to translate *ba(n)tu* (plural of *mu(n)tu* ‘person’) from Kikongo to Portuguese, it was by far *pessoas*, not *gente*, the first word that they could think of.³¹

In a similar fashion to PAL *hende* (< *gente*) in its nominal use, the status of *pessoa* in CP can also work as a generic subject:

³⁰Throughout Portuguese-based Creoles, it seems to be only Batavia Creole that chose the noun *pessoa* from the lexifier: *alung pesua* ‘somebody’.

³¹I learned this the hard way, so to speak: during my first interviews in Cabinda, I included a list of KIK words that speakers had to translate to Portuguese, in order for me to be able to elicit specific phonetic features. *Batu* was on the list; I thought it would be translated with *gente*, which would allow me to register realization of /t/ before palatal vowel in Angolan Portuguese. However, I generally got *pessoa[s]* when I was looking for *gente*.

- (56) *aqui não tem, comboio aqui só [pode] encontrar pessoa nas outra[s] províncias*
 'here [in the province of Cabinda] there isn't any... Trains here [in Angola] you/one can only find in the other provinces (/...people can only find...)'

To complete the parallelism with PAL, it has to be noted that *pessoa* can sometimes adopt an inclusive reading and therefore function as a first person plural expression (*pessoa entende* = 'people [here] understand'/'we understand'). In (57) we also see the active use of a special indefinite (*alguns*) alongside the generic/arbitrary noun *gente* 'people' which, since it is modified by *daqui* 'from here', includes the reference to the speaker too).

- (57) *A: esses congoleses que vêm para cá, eles vêm...eles não falam português, não é?*
 'A: those Congolese who come here, they come... they don't speak Portuguese, do they?'
B: alguns fala[m]
 'B: some do'
A: ah!
 'A: ah!'
B: aprende[m] e sabe[m] falar...Alguns mesmos fala[m] língua deles; como é próximo daqui, pessoa também entende
 'they learn the language and speak... Some speak their language [Lingala] and, since the Congo is close to us, people here can also understand it'
A: aha...e francês também, não é?
 'A: aha! and French they speak too, don't they?'
B: yeah, francês, francês, só que é complicado para a gente daqui
 'B: yeah, French, French also, but French is difficult for the people here'

Finally, *pessoa* can be used as a negative expression in some contexts. In the following example, we observe the alternation between *pessoa* and *ninguém* (the canonical negative indefinite):

- (58) *não pode estar pessoa sem ninguém trabalhá*
 NEG can be anyone without no.one work
 'no one can stay [here] without working [/if no one works]'

The main findings of this section are summarized in Table 5.

Table 5: Indefinite expressions found to date in CP

[– personal]	[affirmative]	<i>uma/alguma/qualquer coisa</i> ~ <i>algo</i> (out-group, polite)	‘something’
		<i>tudo</i>	‘everything’
		<i>muita coisa</i> ~ <i>muita</i> ~ <i>muito</i>	‘a lot’
		<i>pouca coisa</i> ~ <i>pouco</i>	‘few, little’
	[negative]	<i>nada</i>	‘nothing’
		(NEG +) <i>nada</i> ~ (NEG +) <i>uma coisa</i> (counter-expectation)	‘(not...) anything’
[+ personal]	[affirmative]	<i>pessoa</i> ~ <i>uma pessoa</i> ~ <i>alguém</i>	‘someone’
		<i>todo o mundo</i> ~ <i>toda a gente</i> ~ <i>todos</i>	‘everyone’
		<i>muita gente</i>	‘a lot of people’
		<i>pouca gente</i>	‘a few people’
	[negative]	<i>ninguém</i>	‘nobody’
		(NEG +) <i>pessoa</i>	~ (NEG +) <i>ninguém</i> ‘(not...) anyone’

5 Summary and conclusions

The data presented in this chapter are interesting in several ways: on the one hand, I have briefly described three varieties that have been insufficiently studied: Palenquero Creole, the “fossilized” learner varieties (*variedades não nativas*, Gonçalves 2010) of Portuguese as spoken in the Angolan province of Cabinda, and, to a lesser extent, the Kiyombe dialect, which has traditionally been less well described than, for example, Southern Kikongo. On the other hand, I have analyzed a specific grammatical phenomenon from several perspectives, including that of the realization vs. non-realization of the very phenomenon; In other words, I have identified a series of linguistic variables that should be further investigated by future research (for example, *un kusa* vs. *aggú(n) kusa* ‘something’

or *kusa* vs. *ná* ‘nothing’ in Palenquero, *mu(n)tu mosi* vs. *mu(n)tu* ‘someone’ in Kikongo, *muito* vs. *muíta coisa* vs. *muíta* ‘a lot’ in Cabindan Portuguese, etc.).

Nevertheless, I have collected all this data to make a (modest) contribution to the field of contact linguistics. More specifically, the process of creolization has been considered in terms of its relationship to the broader process of second-language acquisition, and both processes have been characterized here regarding the extent to which they are determined by the influence of substrate languages (L1) (cf. Winford 2008, 2012).

My line of reasoning consists of several steps:

- (1) The use of generic-noun-based indefinites reveals itself as a much more frequent and idiomatic strategy in PAL than in its lexifier (Spanish); moreover, some indefinites of this type would surely be impossible in Spanish (starting with the example of *un poko kusa* in the introductory section).
- (2) If we ask ourselves how this structural difference between the Creole and its lexifier emerged, we must pay attention to the three components of creolization: the linguistic universals (during first- and second-language acquisition), the further development of structural tendencies already existing in the lexifier, and the substratal influence. In any case, it is assumed that all three components of creolization always interact to some extent: according to Neumann-Holzschuh & Schneider (2000: 1), “substrates and superstrates appear to offer structural possibilities from which elements of emerging structures are selected on the basis of universal preferences, typological affiliation or formal similarities”. In theory, at least, there is still another possibility, according to which PAL did not develop generic-based-noun indefinites right from the beginning (during its early formative period) but only at a later stage of its history – a linguistic history that has already lasted about four centuries; unfortunately, this line of research is not feasible, since we do not have any written manifestation of PAL until the second half of the 20th century (there is no such thing as “Palenquero historical linguistics” – at least not yet³²). However, it should not go unnoticed that PAL exhibits some grammatical uses of generic nouns that seem to be pretty much idiosyncratic (unknown to both substrate and superstrate, as well as to the universal tendencies of creolization described in the specialized literature) (see examples like (15)). A little speculation: if, perhaps, the Creole developed such grammatical uses of generic nouns on its own, why could it not also have developed some other uses like generic-noun-based indefinites?

³²See, however, the texts analyzed by Gutiérrez Maté (2012).

- (3) Each of the three components of creolization could, to some extent, account for the use of generic-noun-based indefinites in PAL in Creole (we exclude here the fourth kind of explanation outlined above, since there is no way to check it empirically). The universal preference for analytical structures in interlanguages may have triggered the aversion to special indefinites in language contact varieties. The avoidance of irregular morphology (like that of personal indefinite pronominals with the endings *-ie/-ien*) can also be seen as the result of typical interlanguage developments. In addition, it could be assumed that the universals of L1 acquisition favor the generalized use of the canonical phrase structure [DETP Det [NP N (Adj)]], which, again, leads to the overt use of nominal heads (such as *kusa* in *mucho kusa*). As for superstratal influence, it is noteworthy that Colonial Caribbean Spanish (as well as, surely, other varieties of Spanish, which, however, did not play any role in the formation of PAL) occasionally exhibited generic nouns with a meaning close to that of an indefinite pronoun; furthermore, we know other Romance languages that have generalized this type of indefinites (cf. French *quelque chose*, Port. *alguma coisa*, etc.) without apparently being influenced by other languages. Finally, in many Creoles, and most certainly in the one studied here, the substrate may have conditioned the use of generic-noun-based indefinites in the Creole, insofar as this is also the canonical strategy for the formation of indefinite expressions in Kikongo.
- (4) At first sight, it might seem that the very nature of the grammatical phenomenon studied here prevents the isolation of the real effect of substratal influence from the other two components of creolization: after all, generic nouns can easily adopt other semantic values in discourse, including that of indefinite expressions, which already lays the foundation for the linguistic change “generic noun > indefinite pronoun” (with no need to think of language contact). Nevertheless, there are two facts that give rise to the suspicion of substratal influence being the ultimate trigger of the linguistic change analyzed here: first, according to the WALS, only $\frac{1}{4}$ of the world’s languages employ generic nouns as the main strategy for the formation of indefinite expressions; $\frac{1}{4}$ is certainly a not negligible figure, but if Creole language structures were the result of genuinely “creative” universal changes, not conditioned in any way by the contributing languages, it would have been more likely that PAL would have developed the most common type of indefinites in universal terms, i.e. interrogative-based indefinites, which are present in 60% of the world’s languages. Secondly, if

we focus on the contributing languages, it has to be acknowledged that no variety of Spanish seem to make a predominant use of generic nouns (most certainly not to the point of having displaced some special indefinites, as happens in the case of the PAL); that is, if the change “generic noun > indefinite pronoun” were already somehow anticipated in Spanish itself, we would expect that some other Hispanic vernaculars would have spontaneously (without the conditioning of other languages) reached the same result of PAL. However, as far as I know, there is no such Hispanic variety. Finally, it has to be noted that understanding the origins of PAL indefinite expressions as the result of Kikongo language transfer is consistent with the general tendency found throughout Creole languages; according to this, Creoles adopt, to a large extent, their preferred type of indefinite pronouns from substrate languages (so it seems, at least, when comparing the materials of the WALS with those of the APICS).

- (5) The previous point allows us to assume that Kikongo is primarily responsible for the formation of PAL generic-noun-based indefinites (out of Spanish lexical materials). Thus, if we accept this hypothesis as valid, we are in a position to study the “transferability” of Kikongo indefinites also as a measure to compare different types of varieties resulting from the Ibero-Romance/Kikongo language contact with each other and, relying on such comparison, to even set some quantitative and qualitative limits between such varieties (partially restructured vs. Creoles). It is this objective that leads us to take CP into account: as a matter of fact, PAL and CP are one of the very few cases worldwide in which we can find a Creole and a restructured variety that have in common both their substrate and their superstrate (if we accepted that the corresponding dialect continua – respectively, Kikongo and Ibero-Romance – are homogeneous enough as to consider them as typological unities).
- (6) Unlike Spanish, Portuguese has developed towards the “mixed” type of indefinites, so that some generic-based-noun indefinites (like *alguma coisa*) have partially displaced some special indefinites (like *algo*); consequently, the typological unity of the two Ibero-Romance languages is not complete. However, Portuguese still retains some of its special indefinites (*alguém*, *ninguém*, *tudo*, etc.), which is enough to analyze possible restructuring phenomena. As for the particular case of CP, this variety does not exhibit uses of generic-based-noun indefinites that are really unknown in other varieties: even those uses that seem to “diverge” the most from Standard

Portuguese varieties, like the example (38) *Tem coisa para falar?*, can be heard, for instance, in Brazil. At the current state of this investigation, it is unclear whether uses like (38), or the frequent use of *muita coisa*, or the common use of (*uma*) *pessoa* instead of *alguém* are really more frequent in Cabinda than in other lusophone regions; even though the answer to this question may be positive (this is also my first intuition about it), further research is much needed. Be that as it may, there is no doubt that special indefinites (*alguém*, *nada*, *tudo*, etc. and, in some contexts at least, *algo*: cf. 36–37) are also relatively common in CP, even among elderly people. This is a remarkable difference between CP and PAL, which is surely related to the fact that there was far more contact with the superstrate in late-colonial and post-colonial Cabinda than in San Basilio de Palenque at the time of Creole formation (and this difference is independent of the fact that Portuguese, unlike Spanish, already makes general use of the generic-noun-based indefinite meaning ‘something’). Since contact with the non-restructured version of the European language was, to some extent, available, structural simplification (imposed by L2 acquisition) and linguistic “creativity” (during L1 acquisition) played a far less significant role in CP than in PAL. In addition, the language ecology of Cabinda is different from that of SBP for another reason: interlanguages were never “good” from an adaptive and evolutionary perspective (Mufwene 2001) and never became part of a new identity, the kind of “Creole identity” – neither European nor exactly African³³ – that we find in SBP.

³³Today, SBP is considered to be *un chito ri Afrika andi Amerika* (‘a small piece of Africa in America’) by many *Palenqueros* (and by all local tourist guides), but this perception results from a simplification (and a re-ideologization) of the traditional Palenquero identity. It is very doubtful that the founders of SBP, which did not all come directly from Africa, would have had such a perception. If it were so, we would not be able to explain why Kikongo is no longer spoken in the village. In this regard, it should be remembered that the village arose from a gradual coming together of multiple *palenques* (‘maroon communities’), all located in the nearby Sierra de María and/or the neighboring region opposite to the shores of the Magdalena river (Navarrete 2008, Gutiérrez Maté 2016). This gradual process surely extends from the end of the 16th century to the end of the 17th century, that is, from the foundation of the so-called “palenque del Limón” (ca. 1580–1634) to the foundation of the palenque de San Basilio (from the remains of the palenque “San Miguel Arcángel”) or even to the peace treaty between Palenqueros and Hispanic civil authorities (1713–1714). According to this treaty, which I have recently consulted in the Archivo General de Indias, approximately half of the founders of SBP were “negros criollos”, i.e. Black Creoles – born in the mountains, or in the haciendas nearby Cartagena. The socio-identitarian processes that took place in colonial *palenques* were actually very complex, and the formation of SBP was even more idiosyncratic, insofar as Palenque was the first maroon community to obtain its freedom from a colonial administration.

This study lays the foundation for future qualitative and quantitative research on the loss of special indefinites in restructured Ibero-Romance varieties spoken in current or former multilingual scenarios. Future studies will also have to solve some of the structural and variational issues that this work has not discussed. At a structural level, this chapter has not analyzed the distribution of generic-noun-based indefinites with indefinite determiners ('a', 'some', 'any'); however, it could be assumed that this type of indefinites (like PAL *un kusa* 'something') are a compromise (or convergence) between the substrate's preferred strategy for building indefinite pronouns (i.e. bare generic-noun-based indefinites) and a secondary strategy used in Ibero-Romance for the building of indefinite expressions (i.e. generic nouns modified by indefinite determiners). At a variational level, there are still other problems that need to be addressed in the future: as regards PAL, it is unclear how we can distinguish code-switched elements (including special indefinites) from Spanish elements that have been incorporated in the Creole (perhaps even centuries ago); as regards CP, it will be necessary to account for the fact that different sociolinguistic variables (age, gender, level of literacy, time of exposure to Portuguese, etc.) and idiolectal preferences result in different linguistic data. In this regard, the generalizations made throughout this chapter have to be understood within the wider context of a research project that deals with typological change from the perspective of language contact (Gutiérrez Maté in preparation), especially in those cases in which different contact varieties have the same substrate and the same superstrate.

Acknowledgments

I would like to express my gratitude to the people and institutions that helped me obtain the necessary data to show the conclusions of this study. On the one hand, the collection of the data from Cabinda was possible thanks to two fieldwork trips funded by the *Förderprogramm für den wissenschaftlichen Nachwuchs* of the University of Augsburg. On the other hand, obtaining the Palenquero data was possible thanks to Armin Schwegler, who showed his generosity in two aspects: firstly, he granted me access to his initial recordings in Palenque, which are especially valuable today, since they depict traditional varieties of Palenquero (these recordings formed the corpus of my postdoctoral project at the University of California, Irvine, in 2014 and 2015, funded by the the P.R.I.M.E. program of the German Academic Exchange Service and the *European Research Council*); secondly, he facilitated contact with the locals, especially with his friend Víctor Simarra, who was an excellent (and necessary) collaborator available at all

times during the interviews I made in the village in the summer of 2017. Lastly, I would like to thank Abel Massiala, my favorite informant from the Mayombe region and a great collaborator during my research stays in Cabinda, and Maximilian Rieder, a Master's student who is working as a student research assistant (Ger. HiWi) for my project on Angolan Portuguese and has been hired thanks to the Philological-Historical Faculty of the University of Augsburg (through its *Haushaltsmittel für Forschungsvorhaben nach Typ A*).

References

- Almeida, Maria Clotilde. 2008. Youngspeak, subjectification and language change: The case of *bué*. In Maria Clotilde Almeida, Bernd Sieberg & Ana Maria Bernardo (eds.), *Questions on language change: Proceedings of the international colloquium – 16 november 2006 – faculty of letters-Lisbon*, 117–132. Lisboa: Colibri.
- Bickerton, Derek. 1990. *Language and species*. Chicago: University of Chicago Press.
- Bickerton, Derek. 2016 [1981]. *Roots of language*. Berlin: Language Science Press. DOI: 10.17169/langsci.b91.109.
- Biyoko Mabua, Jean-Claude Simon. 2017. *Lexique Kiyombe-Français, Français-Kiyombe*. Saint-Denis: Éditions Connaissances et Savoirs.
- Bontinck, François. 1978. *Le cathéchisme kikongo de 1624*. Réédition critique. Brussels: Koninklijke Academie voor Overzeese Wetenschappen.
- Bosque, Ignacio. 1989. *Las categorías gramaticales: Relaciones y diferencias*. Madrid: Síntesis.
- Bosque, Ignacio & Javier Gutiérrez-Rexach. 2009. *Fundamentos de sintaxis formal*. Madrid: Akal.
- Bostoen, Koen. 2012. *Kikongo dialect continuum: Internal and external classification*. Paper presented at the Niger-Congo Conference, Paris, 18–21 September 2012.
- Bostoen, Koen & Gilles-Maurice de Schryver. 2015. Linguistic innovation, political centralization and economic integration in the Kongo kingdom: Reconstructing the spread of prefix reduction. *Diachronica* 32(2). 139–185. DOI: 10.1075/dia.32.2.01bos.
- Bostoen, Koen & Gilles-Maurice de Schryver. 2018. Seventeenth-century Kikongo is not the ancestor of present-day Kikongo. In Koen Bostoen & Inge Brinkman (eds.), *The Kongo kingdom: The origins, dynamics and cosmopolitan culture of an African polity*, 60–102. Cambridge: Cambridge University Press.
- Cameron-Faulkner, Thea, Elena Lieven & Anna Thiekston. 2007. What part of *no* do children not understand? A usage-based account of multiword negation. *Journal of Child Language* 33. 251–282.

- Campos, Ednalvo Apóstolo & Rosana Siqueira de Carvalho do Vale. 2018. As proformas pronominais ‘esse um’ e ‘aquele um’ e a comunidade quilombola de Jurussaca. *Papia* 28(2). 239–254.
- Cardoso, Hugo C. 2013. Diu Indo-Portuguese structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The atlas of Pidgin and Creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://apics-online.info/contributions/39> (15 October, 2022).
- Carter, Hazel. 1999. *A sketch of Kongo Grammar (Zombo dialect, Angola)*.
- Chaudenson, Robert. 1992. *Des îles, des hommes, des langues*. Paris: L’Harmattan.
- Chaudenson, Robert. 2003. *La créolisation: Theorie, applications, implications*. Paris: L’Harmattan.
- Chicuna, Alexandre Mavungo. 2018. *Portuguesismos nas línguas bantu: Para um dicionário português-kiyombe*. 3rd edn. Lisboa: Colibri.
- DeCamp, David. 1971. Toward a generative analysis of a post-creole speech continuum. In Dell Hymes (ed.), *Pidginization and creolization of languages*, 349–370. Cambridge: Cambridge University Press.
- de Friedemann, Nina S. & Carlos Patiño Rosselli. 1983. *Lengua y sociedad en el Palenque de San Basilio*. Bogotá: Instituto Caro y Cuervo.
- Dereau, Léon. 1957. *Lexique Kikôngo-Français/Français-Kikôngo: D’après le dictionnaire de K. E. Laman*. Namur: Wesmael-Charlier.
- Dom, Sebastian & Koen Bostoen. 2015. Examining variation in the expression of tense/aspect to classify the Kikongo language cluster. *Africana Linguistica* 21. 163–211.
- Gabriel, Christoph, Natascha Müller & Susann Fischer. 2018. *Grundlagen der generativen Syntax: Französisch, Italienisch, Spanisch*. 3rd edn. Berlin: De Gruyter.
- Gonçalves, Perpétua. 2010. *A Génese do Português de Moçambique*. Lisboa: IN-CM.
- Güldemann, Tom. 2010. Sprachraum and geography: Linguistic macro-areas in Africa. In Alfred Lameli, Roland Kehrein & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation*, vol. 2: Language mapping, 561–585. Berlin: Mouton de Gruyter.
- Guthrie, Malcolm. 1967–1971. *Comparative Bantu: An introduction to the comparative linguistics and prehistory of the Bantu languages*. Farnborough: Gregg International.
- Gutiérrez Maté, Miguel. 2012. Lengua afrohispanica, palenquero y español colombiano atlántico en el siglo XVII. Conciencia lingüística y testimonio directo en documentos de archivo. *Revista Internacional de Lingüística Iberoamericana* 10(2). 85–106.

- Gutiérrez Maté, Miguel. 2013. *Pronombres personales sujeto en el español del Caribe: Variación e historia*. Universidad de Valladolid. (Doctoral dissertation). DOI: 10.35376/10324/2517. (30 September, 2020).
- Gutiérrez Maté, Miguel. 2016. Reconstructing the linguistic history of *palenques*: On the nature and relevance of colonial documents. In Armin Schwegler, John McWhorter & Liane Ströbel (eds.), *The Iberian challenge: Creole languages beyond the plantation setting*, 205–229. Madrid/Frankfurt: Iberoamericana/Vervuert.
- Gutiérrez Maté, Miguel. 2017. La partícula focal *jue* (<español *fue*) en el criollo palenquero: ¿Gramaticalización y/o sustrato? *Revista Internacional de Lingüística Iberoamericana* 15(1). 7–46.
- Gutiérrez Maté, Miguel. 2018. El español vernáculo dominicano y la escritura de una carta oficial: El caso de los besinos de asua en 1756. In Joachim Steffen, Harald Thun & Rainer Zaiser (eds.), *Classes populaires, scripturalité, et histoire de la langue: Un bilan interdisciplinaire*, 535–593. Kiel: Westensee-Verlag.
- Gutiérrez Maté, Miguel. 2020. De Palenque a Cabinda: Un paso necesario para los estudios afroiberorrománicos y criollos. In Gabriele Knauer & Ineke Phaf-Rheinberger (eds.), *Mundos caribeños – Caribbean Worlds – Mondes Caribéens*, 105–138. Madrid/Frankfurt: Iberoamericana/Vervuert.
- Gutiérrez Maté, Miguel. In preparation. *ROMANIA BANTU: Transfer-induced restructuring in two Ibero-Romance/Kikongo varieties: Cabindan Portuguese and Palenquero Creole*. University of Augsburg. (Habilitation thesis).
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
- Haspelmath, Martin. 2013. Indefinite pronouns. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The atlas of Pidgin and Creole language structure*, 349–370. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Haspelmath, Martin & The APiCS Consortium. 2013. Indefinite pronouns. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The atlas of Pidgin and Creole language structure online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <https://apics-online.info/parameters/21.chapter.html>.
- Holm, John. 2004. *Languages in Contact: The partial restructuring of vernaculars*. Cambridge: Cambridge University Press.
- Holm, John, Gerardo A. Lorenzino & Heliana R. de Mello. 1999. Diferentes grados de reestructuración en dos lenguas vernáculas: El español caribeño y el portugués brasileño. In Luis A. Ortiz López (ed.), *El Caribe hispánico: Perspectivas lingüísticas actuales*, 43–60. Frankfurt: Vervuert/Iberoamericana.

- Holmberg, Anders, Aarti Nayudu & Michelle Sheenan. 2009. Three partial null-subject languages: A comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica* 63(1). 59–97.
- Hopper, Paul J. 1991. On some principles of grammaticalization. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to grammaticalization*, vol. 1, 17–35. Amsterdam/Philadelphia: John Benjamins.
- Inverno, Liliana Cristina Coragem. 2009. *Contact-induced restructuring of Portuguese morphosyntax in interior Angola: Evidence from Dundo (Lunda Norte)*. Universidade de Coimbra. (Doctoral dissertation).
- Koch, Peter. 1997. Diskurstraditionen: Zu ihrem sprachtheoretischen Status und ihrer Dynamik. In Barbara Frank, Thomas Haye & Doris Tophinke (eds.), *Gattungen mittelalterlicher Schriftlichkeit*, 43–79. Tübingen: Narr.
- Kouwenberg, Silvia. 2013. Papiamentu structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The atlas of Pidgin and Creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/contributions/47> (30 September, 2020).
- Kyala, Miguel Barroso. 2013. *Longoka Kikongo*. Luanda: Mayamba.
- Laman, Karl E. 1912. *Grammar of the Kongo language (Kikongo)*. New York: The Christian Alliance Publication Company.
- Laman, Karl E. 1936. *Dictionnaire Kikongo-Français avec une étude phonétique décrivant les dialectes les plus importants de la langue dite kikongo*. Bruxelles: Librairie Falk fils.
- Lipski, John M. 2020. *Palenquero and Spanish in contact: Exploring the interface*. Amsterdam, Philadelphia: John Benjamins.
- Lopes, Célia Regina dos Santos. 2003. *A inserção de a gente no quadro pronominal do português*. Madrid/Frankfurt: Iberoamericana/Vervuert.
- Maglia, Graciela & Yves Moñino. 2015. *Kondalo pa bibí mejó. Contarlo para vivir mejor. Oratura y oralitura en San Basilio de Palenque*. Bogotá: Editorial Pontificia Universidad Javeriana.
- Maho, Jouni. 1999. *A comparative study of Bantu noun classes*. Gothenburg: Acta Universitatis Gothoburgensis.
- Maurer, Philippe. 1998. El papiamentu de Curazao. In Matthias Perl & Armin Schwegler (eds.), *América negra: Panorámica actual de los estudios lingüísticos sobre variedades hispanas, portuguesas y criollas*, 139–218. Madrid/Frankfurt: Iberoamericana/Vervuert.
- McWhorter, John M. 2011. *Linguistic simplicity and complexity: Why do languages undress?* Berlin/New York: Mouton De Gruyter.

- Meeuwis, Michael. 2013. *Lingala structure dataset*. Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.). <http://apics-online.info/contributions/60> (30 September, 2020).
- Michaelis, Susanne Maria. 2017. *Avoiding bias in comparative creole studies: Stratification by lexifier and substrate*. DOI: 10.5281/zenodo.1255790. (30 September, 2020).
- Mingas, Amélia A. 2000. *Interferência do Kimbundu no português falado em Luanda*. Porto: Campo das Letras.
- Moñino, Yves. 2013. Nuevas luces sobre el uso del plural en palenquero y sus orígenes kikongo. *Visitas al patio* 7. 39–58.
- Moñino, Yves. 2017. Past, present and future of Palenquero Creole. In Armin Schwegler, Bryan Kirschen & Graciela Maglia (eds.), *Orality, identity and resistance in Palenque (Colombia): An interdisciplinary approach*, 15–56. Amsterdam/Philadelphia: John Benjamins.
- Mufwene, Salikoko. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Navarrete, María Cristina. 2008. *San Basilio de Palenque: Memoria y tradición*. Cali: Programa Editorial Universidad del Valle.
- Neumann-Holzschuh, Ingrid & Edgar Schneider. 2000. Introduction: “Degrees of restructuring” in creole languages? In Ingrid Neumann-Holzschuh & Edgar Schneider (eds.), *Degrees of restructuring in Creole languages*, 409–436. Amsterdam/Philadelphia: John Benjamins.
- Roberts, Ian. 2012. Clausal hierarchy. In Laura Brugè, Anna Cardinaletti, Giuliana Giusti, Nicola Munaro & Cecilia Poletto (eds.), *Functional heads: The cartography of syntactic structures*, 351–367. Oxford: Oxford University Press.
- Roche, Jörg. 2013. *Mehrsprachigkeitstheorie: Erwerb – Kognition – Transkulturation – Ökologie*. Tübingen: Narr.
- Schmidt-Radefeldt, Jürgen. 2010. *Zum Projekt eines kontrastiven deutsch-portugiesischen jugendsprachlicher Varietäten*. http://schmidt-radefeldt.de/media/pdf/jsr/Lemmata_und_Korpora.pdf.
- Schuchardt, Hugo. 1888. Beiträge zur Kenntnis des kreolischen Romanisch. I: Allgemeineres über das Negerportugiesische. *Zeitschrift für romanische Philologie* 12. 242–254.
- Schwegler, Armin. 1993. Subject pronouns and person/number in Palenquero. In Francis Byrne & John Holm (eds.), *Atlantic meets Pacific: A global view of pidginization and creolization*, 141–161. Amsterdam: John Benjamins.
- Schwegler, Armin. 1996. *Chi ma nkongo: Lengua y ritos ancestrales en el Palenque de San Basilio (Colombia)*. Madrid/Frankfurt: Vervuert/Iberoamericana.

- Schwegler, Armin. 2002. On the (African) origins of Palenquero subject pronouns. *Diachronica* 19. 273–332.
- Schwegler, Armin. 2007. Bare nouns in Palenquero: A fresh consensus in the making. In Marlyse Baptista & Jacqueline Guéron (eds.), *Noun phrases in creole languages: A multi-faceted approach*, 205–222. Amsterdam: John Benjamins.
- Schwegler, Armin. 2013. Palenquero structure dataset. In Susanne Maria Michaelis, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.), *The atlas of Pidgin and Creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/contributions/48> (30 September, 2020).
- Schwegler, Armin. 2016a. Combining population genetics with historical linguistics: On the African origins of the Latin America Black and Mulatto populations. In Sandro Sessarego & Fernando Tejedo (eds.), *Spanish language and sociolinguistic analysis*, 33–88. Amsterdam: John Benjamins.
- Schwegler, Armin. 2016b. Truth reset: Pragmatics in Palenquero negation. In John McWhorter Armin Schwegler & Liane Ströbel (eds.), *The Iberian challenge: Creole languages beyond the plantation setting*, 231–267. Madrid/Frankfurt: Iberoamericana/Vervuert.
- Schwegler, Armin. 2017. On the African origin(s) of Palenquero. In Armin Schwegler, Bryan Kirschen & Graciela Maglia (eds.), *Orality, identity, and resistance in palenque (Colombia): an interdisciplinary approach*, 51–119. Amsterdam, Philadelphia: John Benjamins. DOI: 10.1075/coll.54.02sch.
- Schwenter, Scott A. 2016. Some issues in negation in Portuguese. In Leo Wetzels, João Costa & Sergio Menuzzi (eds.), *The handbook of Portuguese linguistics*, 425–440. New York: John Wiley & Sons.
- Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics* 10(3). 209–241.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization and genetic linguistics*. Berkeley/Los Angeles: University of California Press.
- van der Wal, Jenneke & Jacky Maniacky. 2015. How “person” got into focus: Grammaticalization of clefts in Lingala and Kikongo areas. *Linguistics* 53. 1–52.
- Winford, Donald. 2008. Processes of Creole formation and related contact-induced language change. *Journal of Language Contact* 2(1). 124–145.
- Winford, Donald. 2012. Creole languages. In Robert I. Binnik (ed.), *The Oxford handbook of tense and aspect*, 428–457. Oxford: Oxford University Press.
- Zanuttini, Raffaella. 1990. Two types of negative markers. *North East Linguistic Society* 20(2). 517–530.

Chapter 5

On the diachrony of Catalan indefinite *qualsevol*

Olga Kellert

University of Göttingen

Andrés Enrique-Arias

University of the Balearic Islands

In this investigation, we aim to characterize the morphosyntactic and semantic properties of Catalan *qualsevol* in diachrony. On the basis of almost 8000 examples extracted from texts encompassing a period from the 13th to the 20th century, we look at variables such as agreement properties, position with respect to the noun, grammatical function, and co-occurrence with strong determiners and quantifiers. Moreover, we analyze the historical evolution of the semantic interpretation of *qualsevol*. This way we are able to trace a grammaticalization path for Catalan *qualsevol* that is similar to the one proposed in the literature for Spanish *cualquiera* (see Company-Company & Pozas-Loyo 2009). More specifically, our analysis reveals that in both languages *qualsevol* and *cualquiera* gradually changed their grammatical functions and semantic interpretations as a result of their origin as a relative clause. Due to the modificational function of the relative clause, it was reanalyzed as a nominal modifier similar to an adjective. When occurring in a prenominal position, the modifier was further reanalyzed as a determiner, whereas the post-nominal one preserved its function as a modifier. While there are many works on similar items in Romance, English and other languages, this is the first systematic analysis of the diachrony of Catalan *qualsevol*.

1 Introduction

The Catalan indefinite *qualsevol* is a Free Choice Item (FCI) similar to Spanish *cualquier(a)*, Italian *qualunque*, French *quelconque* or Romanian *oarecare* (see



Olga Kellert & Andrés Enrique-Arias. 2024. On the diachrony of Catalan indefinite *qualsevol*. In Olga Kellert, Sebastian Lauschus & Malte Rosemeyer (eds.), *Indefinites in Romance and beyond*, 141–175. Berlin: Language Science Press. DOI: 10.5281/zenodo.13759988

Aloni et al. 2010, and references therein). Free Choice indefinites are anti-specific and express referential vagueness. Specificity is given when “a speaker uses an indefinite noun phrase and intends to refer to a particular referent” (von Heusinger 2011: 10). Consider the following example from Spanish with a simple indefinite that refers to a particular referent, namely Dr. Smith:

- (1) Maria se cas-ó con un médico. En concreto con el Dr.
 Maria REF marry-3SG.PAST with a doctor in concrete with the Dr.
 Smith.
 Smith
 ‘Maria married some doctor. Concretely, Dr. Smith.’

Free Choice indefinites such as *cualquier médico* ‘any doctor’ cannot refer to specific individuals (see Alonso-Ovalle & Menéndez-Benito 2010), i.e. *cualquier médico* does not refer to a certain doctor like Dr. Smith. Instead, they convey the meaning that all alternatives count as possible. Thus, the Spanish sentence in (2) conveys the meaning that for every possible book of your consideration you can choose that book.

- (2) Pued-es eleg-ir cualquier libro.
 can-PRS.2SG choose-INF any book
 ‘You can choose any book.’

Like in Spanish, Catalan *qualsevol* in prenominal position has the indefinite modal meaning exemplified in (3a), which has an FCI interpretation (see Alonso-Ovalle & Menéndez-Benito 2010: 29, Rivero 2011). At the same time, this prenominal indefinite is lexically identical to the postnominal one in example (3b). However, the latter does not have the FCI interpretation but rather an evaluative one (EVAL): in example (3b) the speaker qualifies the *home* ‘man’ as ‘unremarkable’ (see Alonso-Ovalle & Menéndez-Benito 2010: 29, Rivero 2011 for this reading in Spanish *cualquiera*):

- (3) a. Pots port-ar=me qualsevol llibre.
 can.PRS.2SG bring-INF=me any book
 ‘You can bring me any book.’ = Every book is a possible option (FCI)
 b. És un home qualsevol
 be.PRS.3SG a man any
 ‘He is an unremarkable man’ = The man is unremarkable (EVAL)

Like Spanish *cualquiera*, Catalan *qualsevol* can be used as a pronoun and as a noun (see (4) and (5) below, respectively). Note that in the noun use, *cualquiera*

and *qualsevol* have an EVAL interpretation, just like the postnominal *cualquiera* and *qualsevol* in (3b). We will show that this difference in interpretation correlates with the +/– lexical category of *qualsevol* as opposed to its grammatical function. Noun and postnominal *qualsevol* is a lexical category, whereas the pronoun and the determiner *qualsevol* are functional/grammatical categories (see §4 and §5).

- (4) a. *Cualquiera* puede hac-er=lo
 b. *Qualsevol* pot f-er=ho
 any can.PRS.3SG do-INF=it
 ‘Anybody can do it’
- (5) a. *Juan* es un *cualquiera*
 b. *Joan* és un *qualsevol*
 John be.PRS.3SG an any
 ‘John is a unremarkable man’

While similar items in Romance, English, and other languages have received considerable attention (see Company-Company & Pozas-Loyo 2009, Rivero 1988 on Spanish, Becker 2014 on Italian and French, Stark 2006 on Italian, Gianollo 2018 on Romance, among others) there is not, as of yet, a systematic description of Catalan *qualsevol*. This chapter aims at filling this gap by providing an empirically supported account of the morphosyntax and semantics of the Catalan Free Choice indefinite *qualsevol* in diachrony. On the basis of some 8000 examples extracted from texts encompassing the 13th to the 20th century, we look at variables such as agreement properties, position with respect to the noun and the verb, grammatical function, and co-occurrence with quantifiers in order to identify the syntactic and semantic status of Catalan *qualsevol*, especially with respect to its stage of grammaticalization, the changes in grammatical function, and the consequences of all of the above for its semantic interpretation.

The chapter is structured as follows: first we begin by providing a short introduction into the historical uses of *qualsevol* and its morphological variants as described in grammars and dictionaries with a brief note on its etymology, including the different hypotheses on its origin (see §2). In §3, we present the corpus and methodology employed in the study. We analyze the corpus data qualitatively in §4. Then we move on to describe the diachronic evolution of *qualsevol* with special attention to its grammaticalization path from a relative clause to nominal modifiers, pronouns and nouns (see §5 and §6). A summary and outlook are presented in §7.

2 Catalan *qualsevol* and its morphological variants

As noted in reference grammars and dictionaries (cf. Alcover & de Borja Moll 1962, Coromines 1985: VI: 889; de Borja Moll 2006: 184), *qualsevol* derives from the combination *QUALE SE VOLET, that is, it is a compound of relative *qual* ‘which’, the impersonal pronoun *se* ‘one’ and third person singular present indicative of *voler* ‘want’. The variant *qualsevulla*, which derives from a present subjunctive of *voler* (*QUALE SE VOLEAT), appears also since medieval times and is still used nowadays, but only in formal written registers.

The traditional plural forms for *qualsevol* and *qualsevulla* are *qualssevol* and *qualssevulla*, respectively, in which the relative *qual* is inflected for number. However, as final -s becomes silent when combined with the initial s- of *se vol*, plural and singular forms are indistinguishable in actual speech and therefore number distinction is just a matter of a spelling convention. In old texts the plural of *qualsevol* may appear written in three separate words: *quals se vol* as in (6a) below. Likewise it is quite common to find examples with no double ss (i.e. *qualsevol* rather than *qualssevol*) accompanying a plural noun phrase; this means that from early on plural inflection for *qualsevol* was only sporadic (6b). Also, by the late 1300s there appears a new analogical plural with -s at the end of the whole sequence: *qualsevols*, as in example (6c) featured in Alcover & de Borja Moll (1962). This new form, which is not allowed in standard normative Catalan, is evidence that by the 14th century speakers had reanalyzed the sequence *qualsevol* as one word. The verb following *vol* is an indicator that at this time it was still compositional.

- (6) a. (*Spill* 431, ca. 1460)
Quantes s-ón vive-s | qual-s se vol s-ien
how.many be-PRS.3PL alive-PL | which-PL REF want be.SBJV.3PL
‘How many are alive, whichever they may be’
- b. (Pere IV, Cròn. 67 (ca. 1383))
No contrastant-s qualsevol-Ø privilegi-s
not withstanding-PL any-Ø privilege-PL
‘Not withstanding any privileges’
- c. (Hist. Sóller, II, 23 (1370))
Per qualsevol-s crim-s e excesso-s
for any-PL crime-PL and excess-PL
‘For any crimes and excesses’

As for gender, in principle the standard feminine form for *qualsevol* is identical to the masculine form, as the relative *qual* has no distinct feminine form. In dialectal non-standard speech, however, there exists the feminine form *qualsevola* (cf. Alcover & de Borja Moll 1962), which again evidences that the erstwhile compound is treated by speakers as one word. In the next section, we discuss different hypotheses about the origin of *qualsevol*.

3 Origin of *qualsevol*

There is some disagreement in the literature regarding the origin of Romance indefinites such as *qualsevol* (see Company-Company & Pozas-Loyo 2009: 1068–1088 and references therein for a summary and discussion centered on Spanish indefinites). According to some linguists, Romance FCIs represent the direct continuation or the adaptation via calque of Latin FCI compounds *quivis*, *quilibet*, *qualislibet*, and the like, a scenario that we will call the *continuation hypothesis*. For instance, Menéndez Pidal (1928) assumes that *qual quier*, *qui quier*, *qual-sequiera*, and so forth represent the Old Spanish equivalents of the Latin *quilibet*, *qualis-libet*, etc., that is, indefinite relatives compounded of a pronoun and an impersonal verb. According to Meyer-Lübke (1899: 57), Italian *qualunque* is derived from Latin *qualiscumque*, composed from *qualis* ‘which’ + *cumque* ‘ever’ (*qualiscumque* > *qual[is]-unqua[m]* > Old and Modern Italian *qualunque*). As for French, etymological dictionaries assume that the indefinite pronoun *quiconque* has its source in *qui que* + *onques*, which was influenced by Latin *quicumque* (see Becker 2014 citing Godefroy 2006 [=1880–1902]: vol. 6, 511; Bloch & von Wartburg 1975: 525; Gamillscheg 1969: 737; Greimas 1998: 489; Tobler & Lommatzsch 1925–2002: vol. 8, 91).

According to an alternative hypothesis, (i.e. the *grammaticalization hypothesis*), FCIs such as Spanish *cualquiera* are a new Romance structure that emerged as the result of the evolution of relative clauses containing a verb of volition. The authors supporting this hypothesis argue for a grammaticalization process in which relative clauses were reanalyzed as indefinite noun phrases (Palomo 1934, Rivero 1988, Haspelmath 1997, Girón Alconchel 2012, Brucart 1999: §7.5.7, Company-Company & Pozas-Loyo 2009, among others). However, there is no consensus about how exactly the grammaticalization of *cualquiera* has taken place, the reason being that the indefinite *cualquiera* was already documented in the earliest Old Spanish documents. The hypothesis of the grammaticalization of *cualquiera* is thus (just) a hypothesis for which one can find good arguments (at best), but not proofs. According to one suggestion of the spell out of the grammaticalization path (cf. Figure 1, based on Company-Company & Pozas-Loyo 2009:

1086, Spanish *cualquiera* starts out as a free relative clause introduced by the wh-element *cual* ‘which’ and an NP + the volitional verb *quiera* ‘want.3SG.SBJV’¹ (see step 1). The second step is an adjacency structure between the wh-element and the verb *quiera* (see step 2). The authors assume that the adjacency structure is [*qual quier NP que...*] in which *qual* is separated from the noun phrase NP. The adjacency and frequency of *qual quier* has the effect that this sequence becomes reanalyzed and lexicalized as one word, which is no longer perceived as clausal. The indefinite acts as an argument of the main verb *haga* (step 3). The biclausal structure (i.e. the main sentence and the free relative clause) at step 1 is reanalyzed as monoclausal at step 3.

- a. *hag-a en él cual castigo quier-a* (Step 1)
do-IMP.3SG in him which punishment want-SBJV.3SG.
‘do upon him the punishment which you may want’
- b. *haga en él cual quiera castigo* → Free Relative with NP in situ (Step 2)
- c. *haga en él cualquier(a) castigo* → Postverbal Indefinite, monoclausal analysis (Step 3)
‘Do to him whatever punishment you want’

Figure 1: Grammaticalization path of *cualquier* in Spanish (Company-Company & Pozas-Loyo 2009: 1086)

The question now is whether the evolution of *qualsevol* shows signs of grammaticalization. As with *cualquiera*, our ability to answer this question is limited by the lack of direct documentation of spoken Latin and proto-Romance. By the time Catalan starts to be written consistently in the 13th century, the grammaticalization of *qualsevol* as a one word compound is fairly advanced. It is not easy then to ascertain whether the Free Choice meaning of Latin indefinites has continued in Catalan with the mere substitution of the verbal component (thus Latin *-vis* or *-libet* would have been replaced by Catalan *se vol* or *se vulla*), or conversely these indefinites derive from a wider sentence structure, namely a relative clause of the type *en qual lloc se vulla* ‘in any place he wants’, which evolved to *en qualsevulla lloc*. However, based on the data available to us in descriptive studies of medieval Catalan and etymological dictionaries (cf. Alcover & de Borja Moll 1962; Batlle et al. 2016: 552; de Borja Moll 2006: 184; Coromines 1985: VI: 889) we are more inclined to accept the Romance innovation scenario. Indeed *qual se vol*-type compounds in early texts exhibit at least three features that are congruent with those of relative clauses:

¹The mood of the basic form for the derivation of *cualquier* (whether it was subjunctive *quiera* or indicative *quiere*) is also controversial (see Pato 2012).

First, there are numerous examples in which there is no adjacency between the relative and the verbal component of the compound. In these cases, there is virtually no way to tell the difference between a compound and a relative clause, as in example (7a) below (Batlle et al. 2016: 552; Badia i Margarit 2004: 226). Second, in the medieval data the verb *voler* inflects according to the tense and modality of the clause (cf. Coromines 1985: VI:889); therefore we may find imperfect *volia* as in (7b) or even future *volrà* as in (7c). And third, the compounds have different forms depending on features of the antecedent. While the *qual*-compounds (*qualsevol*, *qualsevulla*, *qualsequer*, *qualsequira*) can be used with any antecedent, the forms with *que*- (*quesvol*, *quesvulla*) are restricted to inanimate antecedents and the *qui*- form (*quisvulla*) is used with human antecedents (cf. example (7d) from Alcover & de Borja Moll 1962). Only the universal compounds *qualsevol* and *qualsevulla* have had continuity into modern Catalan.

- (7) a. pot f-er e elég-er qual demanda=s
can.PRS.3SG make-INF and choose-INF which request=REF
vol
want.PRS.3SG
'he can make and choose any request he wants'
- b. arremir-en junte-s dos cavaller-s sarraïn-s a dos
challenge-PST.3PL together-PL two knight-PL Moorish-PL to two
altre-s nostre-s, qual-s se vol-ia de la ost
other-PL our-PL, whichSPL REF want-PST.3SG of the army
'two Moorish knights challenged two other people among us,
whoever they were, of our army'
- c. altre qual-se-vol-rà que faç-a les-
another, which-REF-want-FUT.3SG that do-SBJV.3SG the-PL
citacion-s
summon-PL
'another one, whoever it will be who will do the summons'
- d. Senyor caualler, qui-s-vull-a siau, Tirant
Sir knight, who-REF-want-SBJV.3SG be.SBJV.2SG, Tirant
'Sir knight, whoever you may be, Tirant'
Tirant lo Blanc, c. 60 (1490)

The preceding examples are evidence that the *qualsevol*-type compounds derive from a sentence structure rather than a phrasal compound. If Catalan *qualsevol* had emerged as a mere calque or replacement of *quivis* and *quilibet*, it would

have been a cohesive compound from the beginning, which typically would involve certain restrictions, such as strict adjacency of its elements and lack of verb inflection motivated by elements in the sentence (see Company-Company & Pozas-Loyo 2009: 1113 for a similar argument concerning Spanish indefinites).

While, as we have said, the available data does not allow to trace the remote structural origins of these elements, we consider that the evidence in favor of a sentence level origin for *qualsevol* type compounds, rather than a calque or adaptation of Latin compounds, is compelling. It is also true, however, that, from the earliest texts, the vast majority of occurrences of *qualsevol* and *qualsevulla* already exhibit the properties of a cohesive phrasal compound, such as a preference to be written together or the frequent loss of inflection for number.

In the pages to follow, we aim to provide an empirically supported account for the development of these phrasal compounds in diachrony. We look at the historical evolution of different properties, such as number agreement, syntactic function and position of *qualsevol*, in order to trace the grammaticalization path of these compounds from relative clauses to indefinites with an FCI meaning. Moreover, we will look at the different interpretations of indefinites such as the evaluative meaning, in order to see how change in meaning is correlated to the grammaticalization of indefinites such as *qualsevol*.

We will restrict our investigation primarily to *qualsevol*-type forms, as this is the compound that has a continuation into modern Catalan and constitutes the overwhelming majority of FCIs in the history of the language. Therefore, we will not trace the evolution of less frequent compounds such as *quesvol*, *quesvulla*, *quisvulla* and the like.

4 Corpus and Methodology

As we write this chapter there is only one publicly available historical corpus of Catalan, the Corpus Informatitzat del Català Antic (CICA) online at <http://cica.cat/index.php>. This corpus contains texts from ca. 1200–1599 for a total of 6.8 million words, with just a few texts from the 1600s. For later periods in the history of the language we have used two more corpora through personal communication. The first one, the Corpus Informatitzat de la Gramàtica del Català Modern (CIGCMo; Antolí Martínez 2018) encompasses texts produced between 1600–1832 for a total of 5.5 million words. The second one is the Corpus Textual Informatitzat de la Llengua catalana (CTILC) which covers the latest period, from 1833–2003, and with 82 million words is significantly larger than the two other corpora. As there is some overlap between CICA and CIGCM, we have checked carefully to make sure we eliminated all repeated examples. By combining these

three corpora we were able to get data for the whole written history of the language.

We searched the corpora for all occurrences of the relevant forms of the free-choice indefinites: singular forms *qualsevol*, *qualsevulla*, as well as *qualsevolgués* and *qualsequer*, formed with the preterite of *voler* and the verb *querir*, respectively; and plural forms *qualssevol*, *qualssevulla* and *qualsevols*. As we are not concerned with spelling or phonetic variation we subsumed all graphic and dialectal variants (*quansevol*, *colsevol*, *cualsebol*, *cualsevol*, and many others) under their corresponding normalized form. Likewise, in our intention of tracing the evolution of grammaticalized indefinites, we have limited our searches to one word compounds, and as such we have not conducted searches for the separate elements that may occur in the compound. Likewise, we left aside the handful of examples of non-universal indefinites, which only allow for either human antecedents, such as *quisvulla*, or inanimate ones, such as *quesvol*.

As expected in any large scale investigation that is primarily based on historical corpora we had to deal with some common methodological issues (cf. Enrique-Arias 2012, 2009). For starters, we face the problem that not all historical periods of the Catalan language are represented equally. For instance, in the CICA corpus there are less than one million words for the 13th century, as opposed to more than two million for the 14th century and over three million for the 15th, the reason being that the widespread use of written Catalan starts late in the 13th century. As a result, in our combined corpus there were only 25 examples of FCIs for the 13th century compared to several hundreds for the subsequent medieval centuries. At the same time, there were almost two thousand examples for the 19th century and more than 17,000 for the 20th century. This disparity in numbers is also related to the diverse size and scope of the different corpora and, more specifically, the disproportionate number of data for the 20th century in the CTILC compared to the other centuries. To make sure that we do not overlook any important data, especially in the earlier centuries where numbers are relatively lower, we have decided to analyze all occurrences in the corpus except for the 20th century where we have limited our analysis to a randomized sample of 2000 occurrences of *qualsevol* plus 109 examples of *qualsevulla*, which is the corresponding proportional share of this form relative to *qualsevol* for this century. Table 1 features the total number of tokens per century in our database.

Table 1: Number of tokens in the database sorted by century

Century	13	14	15	16	17	18	19	20	Total
Tokens	25	367	696	595	1011	1182	1842	2109	7827

Another problem in corpus based linguistic research is that of the distribution of registers, genres and dialects. During the period known as *The Decadence*, Catalan became less used in cultural contexts: as a result, the portion of the corpus for the 16th to 18th centuries has very few literary works and is primarily made up of notary documents and personal diaries. We are aware that this could have an impact on the results, as the creation and evolution of indefinite compounds is greatly determined by textual genre (cf. Company-Company & Pozas-Loyo 2009: 1107). The same could be said of dialectal variation, as certain forms, such as *qualsevulla*, may be associated with specific dialects. In future research it will be interesting to control for these variables by incorporating a more fine grained analysis of the data and with better control for genre and dialectal variation.

A final issue is that the editors of the CICA have normalized the medieval spellings and in doing so they have unified word separation. For instance, *quals se vol sien* 'whichever they may be' in Jaume Roig's *Espill* (cf. Alcover & de Borja Moll 1962: s.v. *qualsevol*) is rendered *qualssevol sien* on the CICA corpus. This type of normalization eliminates potentially interesting information concerning the grammaticalization of the compound, as the orthographic conventions used by scribes, that is, whether different forms are written separated or bound, and whether or not any constituents can intervene between them, are common criteria to determine the degree of fusion of the elements that take part in a compound. At any rate, as we will primarily be concerned with the evolution of *qualsevol* once the compound is already set, this problem affects just a handful of examples.

Despite these problems we are confident that we have obtained the best historical corpus data available for Catalan. In total we have taken into account 7829 tokens of *qualsevol*-type compounds, which, we consider, is a rather robust data base to extract some generalizations on the evolution of this structure.

Once we had extracted the examples, we coded for those factors that indicate relevant functional and semantic changes in the historical evolution of FCI and that allow us to trace the grammaticalization path of the compound. As we have already pointed out, there is no comprehensive account of the evolution of Catalan *qualsevol*; therefore, in selecting the factors to be analyzed, we need to rely on previous studies for Spanish and other languages, as well as generalizations stemming from grammaticalization theory.

The first factor that we consider is allomorphic variation. As already explained, because *qualsevol*-type structures originate in a relative clause, the medieval Catalan compounds may exhibit allomorphy according to features of the clause, like tense (past *qualsevolgués*, future *qualsevolrà*) or modality (subjunctive *qualsevulla*), or features of the antecedent, such as animacy (*quisvol* with human antecedents, *quesvol* with inanimate ones, and *qualsevol* with either one). The

reduction in the number of allomorphs is an indicator of the transition from a relative clause structure to a cohesive compound.

Second, we analyze number agreement considering two parameters: whether plural contexts trigger the presence of an overt plural marker and, if this is the case, what the morphological exponent of plural is. As we have explained before, it is quite common to find examples with no double *ss* (i.e. *qualsevol* rather than *qualssevol* accompanying a plural noun phrase); this means that from early on, plural inflection for *qualssevol* was only sporadic. In plural contexts with no inflection for plural (see example 6b above) and when there is a plural marker we have several possibilities: on the one hand, there are the traditional plural forms *qualssevol* and *qualssevulla*, in which the relative *qual* is inflected for number, but there is also the innovative plural form *qualsevols* that features the plural *-s* suffix at the end of the whole compound. The continuation of plural *qualssevol* indicates that in some way speakers still analyze the compound as a combination of several distinct elements: *quals se vol*. On the other hand the emergence of a new plural form *qualsevols* is an indicator that speakers understand the compound as one word. There are at least two other features that are related to the degree of integration of the compound: the orthographic convention used by the scribe (whether the elements in the compound are written separated or bound) and interposition (whether or not any constituents can intervene between the relative wh-element and the verb). However, as the configuration of the texts in the corpus and the search engine do not allow this kind of investigation we will not consider these variables.

Next, we consider the position with respect to the noun, that is, whether *qualsevol* precedes or follows the noun it modifies, as in examples (8a) and (8b), respectively, which we labeled as PRE (prenominal) or POST (postnominal). The reason why we looked at the position with respect to the noun is because we wanted to see the frequency distribution of *qualsevol* as a modifier and whether the frequency is the same across periods in the prenominal and postnominal case. Our working hypothesis is that prenominal modifiers are different from postnominal modifiers in syntactic category. The former ones are of the determiner type, the latter ones are of the adjective type (see §5):

- (8) a. renunci-ava a qualsevol pene-s per ell acusade-s a
 renounce-PST.3SG to any penalty-PL by him demanded-PL to
 la dita ciutat
 the said city
 ‘he renounced to any penalties demanded by him to the said city’
 Manual de consells (1378-1379)

- b. que s-íe castig-ad cóm â criminal qualsevol
 that be-SBJV.3SG punish-PTCP like to criminal any
 ‘that he gets punished like any criminal’
 Febrer i Cardona, Antoni: Daniél ô el vertader cúlto de Dèu restablért
 en l’Oriént (1836)

We also consider the syntactic function of the *qualsevol*-type element, that is, whether it works like a modifier, as in the examples (8a–b) above, or as a pronoun, as in (9), or even a noun as in (10).

- (9) Es clar, qualsevol hauria fet igual. (pronoun)
 be.PRS.3SG clear, any have.COND.3SG do-PTCP same
 ‘It’s clear, anybody would have done the same’
 Ruyra Parada 27. (1919)
- (10) Miris, no=s pens-i que jo sig-a un qualsevol (noun)
 look.IMO, not-REF think-IMP that I be-PRS.SBJV.1SG a any
 ‘look, don’t think I’m a nobody’
 Baró, Teodor: No es or tot lo que llú (1872)

Finally we looked at co-occurrences with other quantifiers and determiners in the same noun phrase, which we labeled as (+ DET), e.g. [determiner *qualsevol* noun] or [determiner noun *qualsevol*]. Moreover, we distinguished between +/– strong quantifiers/determiners (see Zamparelli 2000 on strong vs. weak determiners). The reason behind this parameter is because we wanted to see whether *qualsevol* itself can be analyzed as a determiner or quantifier like ‘some/every’. If it is a determiner, then we expect it not to occur with other strong determiners or quantifiers like ‘every’, ‘none’, ‘some’, as a noun cannot be quantified or determined twice **every none girl* (see Etxeberria & Giannakidou 2014, among others, on double quantifiers). Thus, in a construction like [strong determiner *qualsevol* noun] or [strong determiner noun *qualsevol*], as in (11) below, *qualsevol* cannot be a determiner itself (see 5 on the analysis):

- (11) que mingun ortolà ni ninguna altra *qualsevol* persona per si o
 that no farmer nor no other any person by himself or
 per altri
 by another
 ‘that no farmer nor any other person by himself or through somebody else’
 (CA-MOD 5. Llibre del Mostassaf d’Elx) (1610)

5 Quantitative analyses

5.1 Morphological variants

From the onset, *qualsevol* is the most frequent form in the data. In the 13th century CICA registers 60% of *qualsevol* (15/25) next to 36% *qualsequer* (9/25) and one example (4%) of *qualsevulla*. In the ensuing centuries *qualsevol* will increase its frequency even more to become the only allomorph used in regular speech in modern Catalan. Table 2 summarizes the distribution of the different allomorphs in the corpus:

Table 2: Distribution of the allomorphs of *qualsevol* registered in the database

Century	<i>qualsequer</i>	<i>qualsevolgués</i>	<i>qualsevulla/s</i>	<i>qualsevol/s</i>	Total
13	9	0	1 (4%)	15 (60.0%)	25
14	0	0	19 (5.1%)	348 (94.8%)	367
15	0	4	126 (18.1%)	566 (81.3%)	696
16	0	0	65 (10.9%)	530 (89.1%)	595
17	0	0	22 (2.1%)	989 (97.8%)	1011
18	0	0	9 (0.08%)	1173 (99.2%)	1182
19	0	0	119 (6.5%)	1723 (93.5%)	1842
20	0	0	110 (5.3%)	1999 (94.7%)	2109
Total	9	4	471 (6.2%)	7343 (93.8%)	7827

In the earlier stages of the corpus there are a few cases of *qualsequer* (9 examples or 36%), but we find no examples beyond the 1200s. This could be related to the fact that the verb *querir* ‘look for, want’ became very infrequent in Medieval Catalan and all but disappeared by the 15th century (Coromines 1985: VI: 940). *Qualsevolgués* is also rather infrequent with only four examples in one text, *Tirant lo Blanch*, from 1490. The lack of examples for other forms such as *qualsequira*, *qualsevull* and *qualsevulga* in our corpus confirms that these forms were also rather infrequent.

As for *qualsevulla*, this is the only competitor of *qualsevol* that has a continuous presence in the history of the language. This form, which Brucart (2002: 1551) ascribes to the Valencian dialect of Catalan, experiences an increase in the Middle Ages, from just one example in the 13th century to 5.1% and 18.1% for the 14th and 15th centuries, respectively. After the Middle Ages, *qualsevulla* decreases until it almost disappears in the 18th century. In recent times, *qualsevulla*

has experienced a modest increase, but the form is relegated to formal written registers (Colomina i Castanyer 2002: 566). As for its meaning, it is to all effects an equivalent of *qualsevol* (Alcover & de Borja Moll 1962; Brucart 2002: 1551).

An issue that has been the source of disagreement in the study of indefinites in Spanish is that of the mood of the verbal base, that is, whether *qualquier* derives from subjunctive or indicative. Company-Company & Pozas-Loyo (2009: 120–121) consider that subjunctive was the verbal base because a non-factual meaning is better suited to convey the meaning of indifference and generalization in the indefinite. In the case of Catalan it is clear that both indicative (*qualsevol*, *qualsequer*) and subjunctive (*qualsevulla*, *qualsevulga*, *qualsequiera*), contributed to the creation of the compound, but the form with indicative *qualsevol* was always predominant and, in the end, the only one that continued in spontaneous speech. This outcome is somewhat expected, as third person singular of the present indicative is the most frequent, less marked and more basic form and thus it is the most likely candidate to become fixated once the compound loses autonomy and becomes one word.

In sum, the great variety of allomorphs of FCI formed with combinations of *qual-/qui-/que-* + VERB OF VOLITION in the earlier part of the data constitutes strong evidence of the origins of *qualsevol*-type compounds in a sentence structure in which the verbal element of the compound was inflected in accordance with other components of the sentence. But as grammaticalization set in, the different elements in the compound lost autonomy and the inventory of allomorphs was reduced to the universal indefinite *qualsevol* which is the less marked one: it allows any antecedent, whether animate or inanimate, and the verbal base uses unmarked third person indicative *vol*. As we are about to see, *qualsevol* also lost number agreement which, again, is an indicator of further grammaticalization.

5.2 Plural agreement

With respect to plural agreement we are looking at two parameters: first, whether plural contexts trigger overt number agreement morphology in *qualsevol*, and second, when this is the case, whether we find traditional plural *qualsevols*, in which the *wh*-element of the compound is inflected, or rather we find the new form *qualsevols* which treats the compound as a single word. Regarding the first scenario, it seems that, from early on, there is a good number of plural contexts in which *qualsevol* does not inflect for number, as in examples (12a–b) below where the indefinite exhibits no overt agreement:

- (12) a. A tot-s altre-s qualsevol-Ø contract-es
to all-PL other-PL any-Ø contract-PL
‘to any other contracts’
Reintegració de la Corona de Mallorca a la Corona d’Aragó, Carta 264
(ca. 1300–1349)e
- b. Per qualsevol-Ø person-es estrany-es
For any-Ø people-PL strange-PL
‘For any strangers’
Manual de Consells de la ciutat de València 1 (ca. 1300–1349)

The data contains numerous examples attesting that, from the earliest periods recorded in the corpus, plural agreement for *qualsevol* is rather unsteady (see Table 3). In the 13th and 14th century data less than half (47.5%) of indefinites occurring in plural contexts are inflected for plural.² This percentage gets even lower in the next two centuries (8.8% in the 15th century and 10.2% in the 16th century); this downward tendency, however, is reversed in the following centuries, which exhibit a steady increase in the percentage of forms inflected for number agreement: 19.2%, 60.7%, 85% and 71% for the 17th, 18th, 19th and 20th centuries, respectively (cf. Table 3).

Table 3: Percentage of plural entities with a plural marker

Century	13-14	15	16	17	18	19	20
No agreement	75 53.2%	185 91.2%	132 89.8%	79 73.8%	11 39.3%	16 15.0%	12 28.6%
<i>qualsevols</i>	0 0.0%	3 1.5%	7 4.8%	11 10.3%	17 60.7%	91 85.0%	9 21.4%
<i>qualsevol/-ssevulla</i>	67 47.5%	15 7.3%	8 5.4%	17 15.9%	0 0.0%	0 0.0%	21 50.0%
Total agreement	67 47.5%	18 8.8%	15 10.2%	28 26.2%	17 60.7%	91 85.0%	30 71.4%
Total	141	203	147	107	28	107	42

²As there was only one plural example for the 13th century -which exhibited plural agreement- we have collapsed the data from the 13th and 14th centuries.

If we leave aside the 20th century which, as we discuss below, has its own peculiar evolution, the distribution of number agreement exhibits a V-shape progression: a steady decline during the Middle Ages and a rebound and increase in the Modern Era. These developments are concomitant with, and directly related to, the decline of traditional plural forms *qualssevol* and *qualssevulla* and the emergence and encroachment of the new form *qualsevols*. The older form exhibits a steady decrease to the point of disappearing altogether in the 18th and 19th centuries. At the same time, the new form *qualsevols* appears in the 15th century data (although reference grammars mention examples already in the 1300s) and becomes the only plural form by the 18th century. The loss of *qualssevol* and its replacement with the new plural form evinces that the relative clause origin of *qualsevol*-type forms is not apparent to speakers, who treat the erstwhile compound as a single word. The shift from *qualssevol* to *qualsevols* thus represents a further step in the grammaticalization path of *qualsevol*. Figure 2 summarizes the changes in the distribution of plural agreement forms over time.

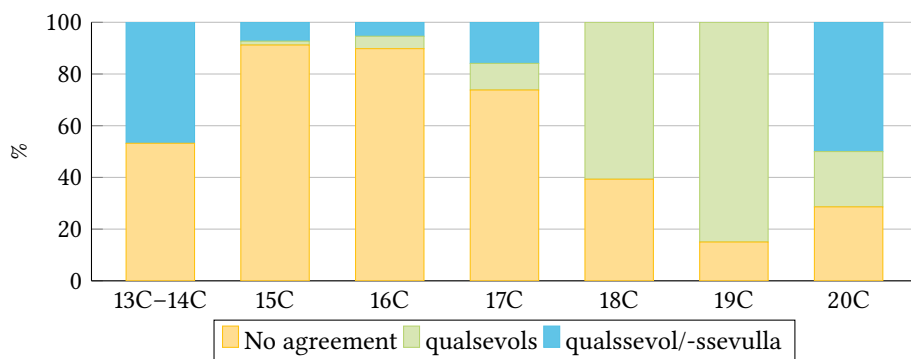


Figure 2: Evolution in the distribution of plural agreement forms for *qualssevol* and *qualsevols*

The developments observed in the 20th century, however, present a stark contrast with the evolution registered in the previous centuries. There is a tremendous decline of the plural form *qualsevols* (as opposed to the upward oriented tendency until the end of the 19th century) and now *qualssevol* and *qualssevulla* reemerge from zero to 50% of the occurrences. This rather unnatural resurgence in the 20th century of forms that had already disappeared in the previous centuries has to do with the written nature of the texts in the corpus, which are obviously affected by the changes in the written conventions for the language. The late 1800s ushered a renewed appreciation of Catalan as a language of culture; following the Decadence period of the 17th, 18th and early 19th centuries,

Catalan first appeared in newspapers and began a gradual entry into universities and scientific academies. At the same time, the Catalan language underwent an unprecedented process of normativization. In particular, Pompeu Fabra's normative works established *qualssevol* as the standard plural form for the literary language, while *qualsevol* was condemned (cf. Coromines 1985: VI: 889). This circumstance explains why in the CTILC corpus non-standard *qualsevol* diminished in the 20th century, while standard *qualssevol* has revived. We must keep in mind that we are dealing with a corpus of written works (literature, newspapers, magazines, essays and scientific and technical materials) that, starting in the early 1900s, are highly influenced by the new normative guidelines. But this recent increase has no repercussion in actual speech, as singular *qualsevol* and plural *qualssevol* are pronounced the same and only differ in the way they are spelled.

5.3 Position and function

The changes explained so far represent a typical grammaticalization path from a Relative Clause to a Nominal Modifier (see Company-Company 2009 for a similar process in Spanish). The new structure, however, has undergone further changes which will be discussed in subsequent sections in detail.

From the earliest periods, *qualsevol* has two basic syntactic functions, that of a pronoun (in the older texts always with a partitive Prepositional Phrase, such as *dels regidors* 'of the councilors' as in (13)), or that of a noun modifier, as in (14) and (15) (see §5.4 below for a more detailed account of the syntactic status of prenominal *qualsevol*).

- (13) *qualsevol dels regidor-s o principal-s de la ciutat*
any of.the councilor-PL or principal-PL of the city
'Any of-the councilors or principals of the city'
Corbatxo - page 67, line: 14 (1397)
- (14) *Ving-a qualsevol temptació,*
come-SBJV.3SG any temptation
'Any temptation may come'
Llull, Blanquerna 6, 7 (ca. 1283)
- (15) *Per qualsevol debilitació del cors,*
by any weakening of.the body
'By any weakening of the body'
Metge Somni I. (1399)

The distribution of the two functions remains fairly stable throughout most of the historical periods in the corpus. Between the 13th and the 18th centuries, the pronoun constitutes approximately 12% to 15% of the total, while the modifier hovers around 85%–82%. The percentage of pronouns goes up, however, to 29.7% in the 19th century, to then lower to 17.6% in the 20th century.

As summarized in Table 4, the modifier function is thus numerically dominant since the beginning and throughout all the periods in the history of the language up to contemporary times.

Table 4: Percentage of the distribution of modifier and pronoun function in the corpus

Century	13-14	15	16	17	18	19	20	Total
Modifier	85.3%	84.8%	87.2%	85.2%	88.2%	70.3%	82.4%	81.5%
Pronoun	14.7%	15.2%	12.8%	14.8%	11.8%	29.7%	17.6%	18.5%

When *qualsevol* is used as a pronoun with no nominal antecedent and without a partitive PP as in (9) above, which we reproduce here as (16) for convenience, it has the meaning of ‘any person’, ‘anybody’:

- (16) Es clar, qualsevol hau-ria fet igual,
be.PRS.3SG clear any have-COND.3SG do-PTCP same
‘It’s clear, anybody would have done the same’
Ruyra Parada 27. (1919)

As illustrated in Table 5 this pronominal use as in (16) already exists in the early texts in the corpus, but only with a few isolated examples; starting in the 16th century, there is a slow but steady increase in the number of cases of the pronoun *qualsevol* with no antecedent and with no partitive PP. The pronominal function grows considerably in the 19th century:

Table 5: Pronoun *qualsevol* with no antecedent and no partitive PP

Century	13-14	15	16	17	18	19	20
N	1	1	5	10	26	159	65
%	0.2%	0.1%	0.8%	1.0%	2.2%	8.6%	3.1%

As for the position of *qualsevol* as a noun modifier, from the beginning, it tended to appear in a predominantly pronominal (PRE) position where it had a

Free Choice meaning, as in examples (14–15). During the early period that encompasses the 14th–16th centuries, however, it is possible to find a few postnominal occurrences of *qualsevol* after bare nouns (POST) which are conjoined with other nouns as in (17). In this early period, postnominal *qualsevol* has only a Free Choice Interpretation as the prenominal *qualsevol*:

- (17) les ciutat-s, castell-s, terre-s e loch-s, baron-s, vasall-s e
 the city-PL, castle-PL, land-PL and place-PL, baron-PL, vasal-PL and
 súbdit-s *qualsevol*
 subject-PL any
 ‘any cities, castles, lands, places, barons, vassals, and subjects’
 Documents de la Cancelleria d’Alfons el Magnànim (15th century)

The first example of POST *qualsevol* preceded by the indefinite determiner *un* is found in the first decades of the 17th century, as shown in (18):

- (18) com si f-os un mort *qualsevol* de cascuna església
 as if be-PST.SBJV.3SG a dead any of each church
 ‘As if it were any dead person from each church’
 CA-MOD 120. Dietari de Pere Joan Porcar-I (ca. 1600-1622)

As stated before, postponed *qualsevol* is very rare in the early stages of the data. As illustrated in Table 6, this situation changed noticeably, starting in the 19th century and continuing into the 20th century data. In the last two centuries, postponed modifiers went from being sporadic to suddenly becoming a sizable proportion of near 10% of the total occurrences of *qualsevol* in its noun modifier function.

Table 6: Frequency of postponed modifier *qualsevol* as opposed to preposed

Century	13-14	15	16	17	18	19	20
N	4	5	2	3	2	114	201
%	1.2%	0.8%	0.4%	0.3%	0.2%	8.8%	11.6%

Another development that happens in the 19th century is the emergence of *qualsevol* as a noun, preceded by an indefinite article. In this new function, *un/una qualsevol* refers to a person of low moral or social status, as illustrated in (19a–b). The data exhibits no examples of this use prior to the mid 19th century, for which we find 13 examples, followed by 16 examples in the 20th century:

- (19) a. Prefer-ia que ella pass-és per una qualsevol,
prefer-PST.3SG that she pass-PST.SBJV.3SG for an any
'he'd rather make her look like a low-class woman'
Oller Febre, I, 154 (1890)
- b. el Clavell és un nuvi de-pega, un titella, un qualsevol
the Clavell be.PRS.3SG a boyfriend fake a puppet a any
'Clavell is a fake boyfriend, a puppet, a worthless man'
Xavier Benguerel: El casament de la Xela (1937)

This new function of *qualsevol* as a noun with the evaluative meaning of 'low class' co-occurs with two different linguistic properties (see Francia & Kellert 2024 [this volume]). First, it correlates with the verbal mood and aspect, i.e. the verb needs to be a predicative verb like 'look like' or 'to be' and it needs to be in indicative present or past tense (see the verbs *passar per* 'pass for' in (19a) and *és* 'to be' in (19b)). Second, *un/a qualsevol* needs to refer to a person. If these two linguistic properties are not present, *un/a qualsevol* is not interpreted as a noun with an evaluative function, but as an elliptical construction [*un/a N qualsevol*]. In this case, the noun rather refers to an entity, not necessarily a person, that was mentioned previously in the discourse. In the following example, *una (altre) qualsevol* refers anaphorically to *una creu* 'a cross':

- (20) no mou á la ànima la contemplació d'una creu gòtica
not move.PRS.3SG to the soul the contemplation of.a cross Gothic
que la de una altre qualsevol de les que ara s=estil-en!
than the of an other any of the that now REF=be.in.style-PRS.3PL
'The contemplation of a Gothic cross doesn't move the soul like any other
of the ones that are now in style!'
Norbert Font i Sagué, *Datos pera la historia de les creus de pedra de Catalunya* (1894)

In the next subsection, we will look into the grammatical status of postnominal *qualsevol* in more detail, especially with respect to its co-occurrence with other determiners and quantifiers. Recall that the reason behind looking at other determiners is to see whether *qualsevol* itself can be analyzed as a determiner or quantifier.

5.4 *Qualsevol* in co-occurrence with determiners and quantifiers

Qualsevol can co-occur with almost every possible determiner/quantifier³: universal (*tot*), existential (*algú, un*), negational (*ni*), bare noun, *altre* ‘other’ + bare noun (see 21–28).

We find data with *qualsevol* N and universal quantifier *tot* with and without coordination: *Tot (i) qualsevol* N: ‘every N, whatever property/identity/kind N might have’. In example (21), *qualsevol* has a different status than in (22), as it is not coordinated with the universal quantifier *tot*:

- (21) al for de València y a tot altre qualsevol dret que
 to.the law of Valencia and to all other any right that
 ting-a introdu-hit en son favor.
 have-SBJVG.3SG introduce-PTCP in his favor
 ‘To the law of Valencia and to all of any other rights that he may have
 been introduced in his favor’
 (CA-MOD 120. Dietari de Pere Joan Porcar-I) (1650–1666)
- (22) y man-á á totes y qualsevol-s persone-s tant laique-s com
 and order-PST.3SG to all and any-PL person-PL so lay-PL like
 eclesiàstique-s, secular-s y regular-s
 clergy-PL, secular-PL and regular-PL
 ‘and he ordered all and any people whether lay, clergy, secular, or regular’
 (CA-MOD 120. Dietari de Pere Joan Porcar-I) (1894)

In the following examples in (23–24), *qualsevol* is used with negative determiners and quantifiers as in *Ni/Ningun N qualsevol*: ‘no/nor N, whatever property/kind/identity N might have’:

- (23) que mingun ortolà ni ninguna altra qualsevol persona per si o
 that no farmer nor no other any person by himself or
 per altri
 by other
 ‘that no farmer nor any other person by himself or through somebody
 else’
 (CA-MOD 5. Llibre del Mostassaf d’Elx) (1610)

³Usually, quantifiers and determiners are analyzed as two distinct categories. Whereas a determiner is a syntactic category represented as the head of the noun phrase (i.e. DP), a quantifier is primarily a semantic category that can be represented syntactically as a (strong) determiner (see Zamparelli 2000, and references therein). However, there are syntactic analyses that assume a syntactic position within the DP for quantifiers, so-called QPs.

- (24) ne per apellació ne per altra qualsevol raó, [...]

neither by appeal nor for other any reason

‘neither by appeal nor for any other reason’

Dietari o Llibre de Jornades, (ca. 1450–1499)

In early periods, postnominal *qualsevol* was used very often with bare nouns ‘other N, whatever property/kind/identity N might have’:⁴

- (25) tirar les dite-s pedre-s, axí corde-s com fusta e altre qualsevol

throw the said-PL stone-PL, like rope-PL as wood and other any

cosa

thing

‘throwing said stones, or strings or wood or any other thing’

Libre del Mostassaf de Mallorca, (ca. 1400–1449)

In later periods, postnominal *qualsevol* was very rarely used with bare nouns as in (26):

- (26) Y crid-e en vá, com dona qualsevulla o un aprenent de

and cry.PST.1SG in vain like woman any or a apprentice of

cuyna?

kitchen?

‘and I yelled in vain, like any woman or a kitchen apprentice?’

Artur Masriera i Colomer, Hamlet príncep de Dinamarca, (1898)

Instead, postnominal *qualsevol* was often used with indefinite determiners such as *Un N qualsevol* ‘some N, whatever property/kind/identity N might have’:

- (27) un xeval qualsevol, sig-a qui sig-a, lo que primé=t

a lad any be-3SG.SBJV who be-3SG.SBJV the what first=dat.2sg

vingu-i á ma

come-SBJV.3SG to hand

‘any lad, whoever it may be, the first thing that comes to hand’

Rossend Arús i Arderiu, Cartas á la dona (1877)
- (28) ab l=excusa de f-er una pregunta qualsevol á la Sra. Pepa

with the=excuse of make-INF a question any to the Ms. Pepa

‘with the excuse of asking Ms Pepa any question’

Narcís Oller, La papallona, (1882)

⁴The semantic interpretation of bare nouns is a controversial topic in the literature. It is standardly assumed that bare nouns are interpreted generically (see Zamparelli 2000 and references therein). We leave the study of bare nouns in Old Catalan for future research.

Table 7 shows the frequency of *qualsevol* (Qlsv) with different types of determiners or quantifiers (represented by the upper number) and the calculated percentages of these frequencies (represented as decimal numbers). This table shows that *qualsevol* co-occurs more often with bare nouns, less often with indefinite nouns, and even less often with universal, existential, and negative quantifiers (> represents the fall in frequency). The hierarchy is schematized in (29).

- (29) bare > indefinite (*un*) > universal quantifiers (*tot*) > negative quantifiers (*ni*) > existential (*algun*)

Postnominal modifier *cualquiera* in structures like [*un/a* Noun *cualquiera*] is often analyzed as an indefinite or quantificational determiner akin to ‘some’ (see Choi & Maribel 2008) or like ‘all/every’ (Aloni et al. 2010) while the status of the indefinite *un/a* in [*un/a* Noun *cualquiera*] is simply ignored.

- (30) [? *un* Noun *hombre* Determiner *cualquiera*]

These analyses are problematic for the *qualsevol* data, given the co-occurrence of postnominal *qualsevol* with indefinite determiners and other quantifiers (see Table 7). As we have already shown, [*Un* N *qualsevol*] and [*Un qualsevol*] rise in frequency from the 19th century, which is also when the evaluative interpretation of [*Un qualsevol*] as ‘unremarkable/low value’ appears. The determiner analysis of *qualsevol* cannot explain the appearance of *qualsevol* as a noun with an evaluative meaning.

Given the problematic analysis of *qualsevol* as a determiner in [*Un* N *qualsevol*] or in [*Un qualsevol*], we would like to suggest a different analysis of *qualsevol* in these configurations. Based on diachronic data (i.e. co-occurrence of *qualsevol* with other quantifiers, as shown in Table 7), we argue in the following §6 that the postnominal as well as the nominal *qualsevol* have the status of a *predicate* with the Free Choice Interpretation in (31) or with the evaluative interpretation in (32) (see also Francia & Kellert (2024 [this volume])):

- (31) *algún/tot/ningún/un/bare* N *qualsevol*
 ‘some/every/none/a/bare N’, ‘whatever identity/property/kind N one wants (literal) or N might have’ FCI
- (32) *un* N *qualsevol*
 ‘some ordinary/low value N’

In the next section, we spell out the diachronic path followed in the evolution of *qualsevol* and answer the question as to how the determiner *qualsevol*, the nominal modifier and the noun *qualsevol* emerged on this path.

Table 7: Distribution of quantifiers, bare nouns, and indefinite nouns with *qualsevol*.

Century	13	14	15	16	17	18	19	20	Total
Bare N Qlsv	5	40	54	58	80	102	247	204	790
Altr	71.43	31.25	34.62	35.58	32.79	25.00	40.49	40.56	35.62
Altr	0	64	48	51	82	202	38	6	491
Qlsv		50.00	30.77	31.29	33.61	49.51	6.24	1.19	22.13
Bare	1	0	1	5	10	26	159	65	265
Qlsv	14.29								
Un N Qlsv	0	0	0	0	1	0	120	220	341
					0.41		19.70	43.74	15.37
Tot Qlsv	0	14	30	25	22	51	32	2	176
	10.85	19.23	15.34	9.02	12.50	5.25	0.40		7.94
Ni Qlsv	0	8	23	25	49	27	14	4	151
	6.25	14.74	15.34	20.08	6.62	2.30	0.80		6.81
Algún Qlsv/ Qlsv Algún	0	2	0	0	0	0	0	2	4
	1.56						0.40		0.18
Total	7	128	156	164	244	408	610	503	2220
100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

6 Diachronic analysis of *qualsevol*

In this section, we will analyze the grammaticalization of *qualsevol*. We will first give a summary of diachronic evidence for the grammaticalization path of *qualsevol* in §6.1 and then provide an account for the reanalysis of *qualsevol* into different grammatical functions in §6.2.

6.1 Summary of diachronic evidence for the grammaticalization of *qualsevol*

Grammaticalization is commonly understood as the process by which a lexical form becomes a grammatical marker, or a grammatical form or construction assumes an even more grammatical function (cf. Kuryłowicz 1965, Lehmann 1982, Hopper & Traugott 2003). There are a number of historical evolutions that have been identified as typical effects of grammaticalization processes, such as the loss of syntactic autonomy, the rigidification of positional patterns, the weakening of referential meaning, phonetic erosion, the reduction of contextual syntactic distribution, and often the change of grammatical status, including the tendency for the grammaticalized form to be integrated into new paradigms (Lehmann 1985, Company-Company 2009).

The historical developments that we have identified so far in the analysis of the evolution of *qualsevol* correspond neatly with the processes that are commonly associated with grammaticalization-type changes. As we have seen, the nominal and verbal components of the indefinite compounds (i.e. the relatives *qual*-, *que*-, *qui*-, and the verb *voler-se*) lost autonomy, since both components stopped being free words to become morphemes of a compound that became a simple word. As such, the relative stopped being inflected for number and thus the plural form *qualssevol* was replaced by a new form *qualsevols*, in which the plural inflection -s was affixed to the end of the verbal component, effectively treating the erst-while compound as a single word. Semantically, the relative clause *qual se vol* lost its compositional meaning of an open proposition with a variable x, as being represented by the wh-pronoun *qual* and the volitional verbal phrase *se vol* [Rel. Cl. *qual se vol*]= ‘one wants x’ (see Caponigro 2014 and Kellert 2015 on semantic interpretation of free relative clauses in synchrony). The new construct (i.e. the one-word-compound *qualsevol*) acquired a new meaning, namely a set of alternatives that is interpreted with respect to a *different* modal verb than the one provided diachronically earlier by the volitional verb. The alternatives of the new construct are interpreted with respect to the modal verb of the *matrix clause* (e.g. *pots portar qualsevol llibre* ‘you can bring any book’) and not with respect

to the volitional modal verb of the relative clause, as it was the case at a prior stage, when the volitional verb was still part of the meaning (as in the relative clause structure *qual se vol*). Clearly, the semantic change whereby indefinites such as *qualsevol* change their interpretation with respect to the new modal verb provided by the matrix verb needs to be worked out in detail in the future (see Kellert 2021).

To summarize, one important trigger for the grammaticalization of *qualsevol* is the fact that the verbal component lost autonomy, since it changed from being inflected for mood and tense in old Catalan (*qualsevolrà*, *qualsevolgués*), to become fixed in the invariable form *qualsevol*. Likewise, there was a loss of autonomy in the pronominal component of *qualsevol* since out of the various existing medieval forms *que*-, *qui*-, *qual*-, only the latter survived. Moreover, the two original forms of the construction changed their categorical status, since both formatives were reinterpreted or reanalyzed as a simple indefinite pronoun: the components stopped being a relative pronoun and a verb, respectively, to become a new form of the Catalan pronominal system, namely the indefinite compound. Finally, there was also a process of paradigmaticization as the new form was integrated into the Catalan paradigm for indefinites along with forms such as *algú* ‘somebody’, *cadascú* ‘each one’, *tothom* ‘everyone’, and so on.

In sum, the evolution of *qualsevol* represents a full grammaticalization path, which does not reflect the mere translation of Latin indefinite compounds as proposed in the Continuation Hypothesis. If that had been the case, we would have found a cohesive compound from the beginning.

6.2 Changes of *qualsevol* into different grammatical functions

We assume that at the very first stage of *qualsevol*, it was analyzed as a Relative Clause (RC) (see Grammaticalization Hypothesis in §3) represented as a complementiser clause (CP) (see Rivero 1988, among others). At this stage, *qual* is a *wh*-element with a *wh*-feature [+wh], which simply marks an element as *wh*-relative or *wh*-interrogative pronoun (see Kellert 2015 on *wh*-features, Kellert 2021.). This pronoun refers to the object argument of the finite verb *vol* inside the finite verbal phrase (represented as TP for Temporal and Finite Phrase). We represent this reference to the object argument by an index *j*:

- (33) [CP *qualj* [+wh] [TP *se vol j*]] First stage=transparent RC
'what(ever) one wants.'

The Relative Clause analysis explains the existence of examples with interposition such as *qual N se vol* in the earlier documents, as in example (7a) repeated here as (34):

- (34) pot fer e eleger qual demanda's vol
 can.PRS.3SG make-INF and choose-INF which request=REF want-PRS.3SG
 'he can make and choose any request he wants'

We analyze [*qual* NP] as a specifier of a free Relative clause (RC) (Kellert 2021):

- (35) [CP [*qual* NP]_j [+WH] [TP *se vol*_j]] First stage= RC
 'whatever NP one wants'

At this stage, *qual se vol* can also modify an overt NP *outside* the Relative Clause or CP, as shown in (36). As *qual* shows plural agreement in the structure in (36), we must assume that, at this point, the relative clause is still transparent, even though the grammaticalization process has already started, as shown by the orthographic representation of *qualsevol* as a single word:

- (36) [Det [NP [ModifP [CP *quals*_j [+WH] [TP *se vol*_j]]]]]
 e.g. *tots deutes qualssevol*
 'all debts whatever kind one wants'

We assume that the relative clause in (36) denotes a *property*, which describes individuals denoted by the noun phrase *deutes* 'debts', as in (37):

- (37) *tots deutes qualssevol*: all x [debts' (x) & qualssevol' (x)]

In this analysis, *qualssevol* has a similar syntactic and semantic status as an adjective with the meaning 'common/ordinary' (see Francia & Kellert (2024 [this volume]), Kellert 2021):

- (38) *tots deutes comuns*: all x [debts' (x) & common' (x)]

The crucial point of our analysis of *qualsevol* as a property in (38) is that this property is assigned to *qualsevol* only at the level when it was reanalyzed as one syntactic category (i.e. a modifier), and not when it still was a relative clause. In other words, the meaning 'common, unremarkable' is part of the diachronic change that arises after the relative clause is no longer perceived as clausal (see Figure 3).

The next step is the loss of RC transparency and the lexicalization of the relative clause into a single compound word. At this stage, *qual* is no longer transparent for plural agreement. The plural agreement is realized instead on the ending of *qualsevol* as in *qualsevol*s. The indefinite is directly interpreted as a nominal modifier without the RC basis:

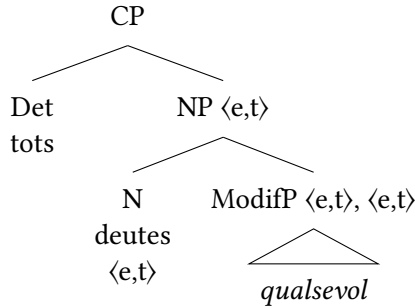


Figure 3: Analysis of *Tots deutes qualssevol*.

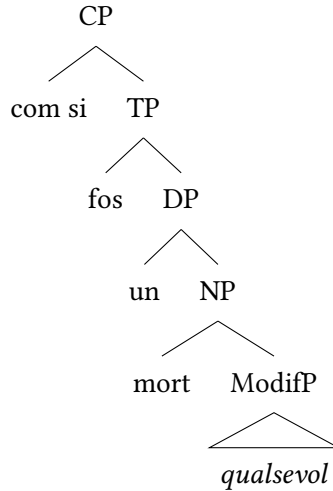
- (39) Modifier *qualssevol* Second stage=loss of RC transparency
- a. cònsols de qualssevol-s viles (16th century)
 consuls of any.PL towns
- b. y universal-s qualssevol-s (17th century)
 and universal.PL any.PL

We assume that bare nouns in Catalan have been replaced by indefinite nouns as represented in (40) (see Lapesa 1975 for this assumption in Old Spanish), i.e. the indefinite determiner *un* in (40b) replaced the (covert/empty) determiner of bare nouns in (40a):

- (40) a. [DET $\emptyset$ [NP N [MODIFP *qualssevol*]]]
 e.g. (42) [...] e súbdits qualssevol
 ‘and subjects whoever they are’
- b. [DET *un* [NP N [MODIFP *qualssevol*]]]
 e.g. (18) [...] *un* mort qualssevol
 ‘any dead person’

In the mid 1800s, *qualssevol* started to appear as a noun as in *una/un qualssevol* ‘a female or male person with low status’. We leave it open as to whether nominalized elements can be interpreted as modifications of covert generic nouns with a gender specification like ‘male person’ and ‘female person’ (see (41a)) or as real nominalizations where *qualssevol* is interpreted as a noun (see (41b)):

- (41) a. [DET *un* [N ‘person’ [+ male]] [MODIFP *qualssevol*]]
 b. [DET *un* [N *qualssevol*]]
 ‘a male person with low status’

Figure 4: Analysis of *Com si fos un mort qualsevol*.

One way to test the two analyses in (41) is using coordination. Under the analysis in (41b), but not in (41a), one should be able to coordinate *qualsevol* with nouns. This will be tested in future research.

Before turning to prenominal *qualsevol*, we will show which data the nominal modifier analysis of *qualsevol* covers so far, and why it is better than previous analyses in the literature (see our review of the literature regarding the analysis of structure (30)). As the postnominal *qualsevol* is not a determiner, it can co-occur with other determiners. It also explains the postnominal position of *qualsevol*, because *qualsevol* has its origin in a relative clause, and relative clauses normally follow nouns. It also explains the adjective-like use of postnominal *qualsevol*. It is a common assumption in the literature that postnominal adjectives in Italian or Romance in general have the syntactic structure of a relative clause (see Cinque 2010). We have shown that postnominal *qualsevol* originates as a relative clause and evolves into a nominal modifier. In that sense, there is a strong parallel between postnominal *qualsevol* and postnominal adjectives. The occurrence of an evaluative meaning is easier to explain under the assumption that *qualsevol* is a nominal modifier rather than a determiner due to its adjective-like and thus lexical status rather than its grammatical status (see Francia & Kellert (2024 [this volume])). A detailed analysis of the different readings of the modifier *qualsevol* and how these readings evolved awaits future research.

The question now is whether the same modifier analysis as suggested in (40) can be applied to the prenominal *qualsevol*. We suggest that the prenominal *qual-*

sevol should be analyzed as a nominal modifier in sentences like in (42), where *qualsevol* is preceded by a determiner and followed by noun:

- (42) al for de València y a tot altre qualsevol dret que ting-a
to.the law of Valencia and to all other any right that have-SBJV.3SG
introdu-hit en son favor.
introduce-PTCP in his favor
‘To the law of Valencia and to all of any other rights that he may have
been introduced in his favor’
(CA-MOD 120. Dietari de Pere Joan Porcar-I) (1650–1666)

The DP *tot altre qualsevol dret* in (42) is analyzed in (43):

- (43) [DET *tot* [*altre* [MODIFP *qualsevol* [NP *dret*]]]]

However, in examples without any overt determiner like *tot* in (42), the prenominal *qualsevol* can be analyzed as a determiner-like attributive element, as demonstrated in (44).

- (44) [DET *qualsevol* [NP *penes*]]
[...] a qualsevol penes [...] (cf. (8a)).
‘to any penalties’

We suggest a similar analysis of the determiner *qualsevol* as shown in (9), which we reproduce here as (45), for pronoun uses of *qualsevol*:

- (45) Es clar, qualsevol hau-ria f-et igual. (pronoun)
be.PRS.3SG clear any have-COND.3SG do-ptcp same
‘It’s clear, anybody would have done the same’
Ruyra Parada 27. (1919)

As for the pronoun use, the noun is analyzed as a generic noun with animate feature with the semantic interpretation of a ‘person’:

- (46) [DET *qualsevol* [NP ‘person’]] hauria fet igual.
‘anybody would have done the same.’

In Figure 5 we summarize what we have shown in this section. The element *qualsevol* originated as a relative clause, then it lexicalized into one word; then, depending on the prenominal or postnominal position, this new one word category was either reanalyzed as a lexical category (i.e. as a postnominal modifier or as a noun), or as a grammatical category (i.e. as a determiner or pronoun). Only the latter development can be defined as a process of grammaticalization in the sense of Lehmann (1985).

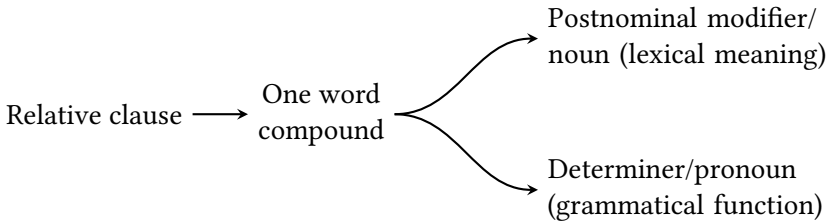


Figure 5: Evolution of *qualsevol*

7 Summary and outlook

In this chapter, we have examined the morphosyntactic and semantic properties of Catalan *qualsevol* in diachrony and we have proposed a grammaticalization path for this structure. We assumed that it started out as a relative clause and that due to its modificational function, it was reanalyzed as a nominal modifier similar to an adjective. The prenominal modifier was further reanalyzed as a determiner, whereas the postnominal one preserved its function as a modifier.

This chapter does not provide any detailed semantic analysis of Free Choice and the evaluative ‘unremarkable’ interpretation and how these two readings are interrelated (see Kellert 2021). In future research, the syntactic functions of *qualsevol* should be examined using contemporary oral data in order to see whether it has grammaticalized any further in Modern Catalan. Finally, it will be important to check in future investigations whether the changes in the grammaticalization path demonstrated for Catalan *qualsevol* coincide with the development of other FCIs in Romance, such as Spanish *cualquiera*, Italian *qualunque/qualsiasi/qualsivoglia*, and French *quelconque*.

Acknowledgements

Olga Kellert is thankful for the funding support from Deutsche Forschungs Gemeinschaft (DFG 256240798) for the project on quantification in Old Italian and other Romance languages and from the Habilitation grant from the Philosophical Faculty of the Georg-August-University of Göttingen. Andrés Enrique-Arias thankfully acknowledges funding from MICIU/AEI (10.13039/501100011033) FEDER, UE for the grants PID2020-116863GB-I00 and PID2023-150917NB-I00. Thanks are also due to Manuel Pérez Saldanya and Jordi Antolí Martínez for their assistance in obtaining data from the Corpus Informatitzat de la Gramàtica del Català Modern (CIGCMod) and the Corpus Textual Informatitzat de la Llengua

catalana (CTILC), respectively. Likewise, we are very grateful to the audience of the workshop “Indefinites in Romance. The limits of an unstable category” at the 36th Romanistentag in Kassel for their helpful feedback. We thank the editors and anonymous reviewers for their thoughtful comments.

References

- Alcover, Joan & Francesc de Borja Moll. 1962. *Diccionari català-valencià-balear*. <https://dcvb.iec.cat/>.
- Aloni, Maria, Ana Aguilar Guevara, Angelika Port, Katrin Schulz & Radek Šimik. 2010. *Free choice items as fossils*. Workshop on Indefiniteness Crosslinguistically (DGfS Berlin).
- Alonso-Ovalle, Luis & Paula Menéndez-Benito. 2010. Modal indefinites. *Natural Language Semantics* 18(1). 1–31. DOI: 10.1007/s11050-009-9048-4.
- Antolí Martínez, Jordi M. 2018. El procés de constitució del Corpus Informatitzat de la Gramàtica del Català Modern (CIGCMod): Objectius, criteris i avaluació. *Notandum* 48. 3–20.
- Badia i Margarit, Antonio Maria. 2004. *Moments clau de la història de la llengua catalana*. València: Universitat.
- Batlle, Mar, Joan Martí i Castell, Josep Moran, Ocerinjauregui & Joan Anton Rabella. 2016. *Gramàtica històrica de la llengua catalana*. Barcelona: Publicacions de l'Abadia de Montserrat.
- Becker, Martin. 2014. *Welten in Sprache: Zur Entwicklung der Kategorie „Modus“ in romanischen Sprachen*. Berlin: De Gruyter.
- Bloch, Oscar & Walther von Wartburg. 1975. *Dictionnaire étymologique de la langue française*. Paris: Presses universitaires de France.
- Brucart, Joseph M. 1999. La estructura del sintagma nominal: Las oraciones de relativo. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 395–522. Madrid: Espasa Calpe.
- Brucart, Joseph M. 2002. La quantificació. In Joan Solà, Maria-Rosa Lloret, Joan Mascaró & Manuel Pérez Saldanya (eds.), *Gramàtica del català contemporani*, 1517–1589. Barcelona: Empúries.
- Caponigro, Ivano. 2014. The semantic contribution of wh-words and type shifts: Evidence from free relatives crosslinguistically. In Robert B. Young (ed.), *Proceedings from Semantics and Linguistic Theory (SALT) XIV*, 38–55. Ithaca, NY: CLC Publications, Cornell University.

- Choi, Jinyoung & Romero Maribel. 2008. Rescuing existential free choice items in episodic sentences. In Oliver Bonami & Patricia Cabredo Hofherr (eds.), *Empirical Issues in Syntax and Semantics 7 (CSSP 2007)*, 77–98. <http://www.cssp.cnrs.fr/eiss7/choi-romero-eiss7.pdf>.
- Cinque, Guglielmo. 2010. *The syntax of adjectives: A comparative study*. Cambridge, MA: MIT Press.
- Colomina i Castanyer, Jordi. 2002. Paradigmes flectius de les altres classes nominals. In Joan Mascaró Joan Solà Maria-Rosa Lloret & Manuel Pérez Saldanya (eds.), *Gramàtica del català contemporani*, 539–582. Barcelona: Empúries.
- Company-Company, Concepción. 2009. Parámetros de gramaticalización en los indefinidos compuestos del español. In Fernando Sánchez Miret (ed.), *Románística sin complejos: Homenaje a Carmen Pensado*, 71–103. Bern: Peter Lang.
- Company-Company, Concepción & Julia Pozas-Loyo. 2009. Los indefinidos compuestos y los pronombres genérico-impersonales omne y uno. In Concepción Company-Company (ed.), *Sintaxis histórica de la lengua española (segunda parte: La frase nominal)*, 1073–1219. Mexico City: Fondo de Cultura Económica–Universidad Nacional Autónoma de México.
- Coromines, Joan. 1985. *Diccionari etimològic i complementari de la llengua catalana*. Barcelona: Curial.
- de Borja Moll, Francesc. 2006. *Gramàtica històrica catalana*. València: Universitat de Valencia.
- Enrique-Arias, Andrés. 2009. Some methodological issues in the corpus-based study of morphosyntactic variation: The case of Old Spanish possessives. In Richard J. Whitt (ed.), *Diachronic corpora, genre and language change*, 261–279. Amsterdam: John Benjamins.
- Enrique-Arias, Andrés. 2012. Dos problemas en el uso de corpus diacrónicos del español: Perspectiva y comparabilidad. *Scriptum digital* 1. 85–106.
- Etxeberria, Urtzi & Anastasia Giannakidou. 2014. Anti-specificity and the role of number: The case of Spanish *algún/algunos*. Paper presented at *Rencontres d'Automne de Linguistique formelle : Langage, Langues et Cognition (RALFe 2014)*, Université Paris 8, Paris.
- Francia, Marika & Olga Kellert. 2024. Argentinian Spanish *cualunque* and Italian *qualunque*. In Olga Kellert, Sebastian Lauschus & Malte Rosemeyer (eds.), *Indefinites in Romance and beyond*, 45–86. Berlin: Language Science Press. DOI: 10.5281/zenodo.13759984.
- Gamillscheg, Ernst. 1969. *Etymologisches Wörterbuch der französischen Sprache*. Heidelberg: Winter.
- Gianollo, Chiara. 2018. *Indefinites between Latin and Romance*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780198812661.001.0001.

- Girón Alconchel, José Luis. 2012. Gramaticalización como creación de lengua a partir del habla: Relativos e indefinidos compuestos en los Fueros de Aragón y en el Fuero de Teruel. *Archivo de Filología Aragonesa* 68. 15–38.
- Godefroy, Frederic. 2006 [=1880–1902]. *Dictionnaire de l'ancienne langue française et de tous ses dialectes du 9e au 15e siècle*. Paris: Editions Champion Electrique.
- Greimas, Algirdas Julien. 1998. *Dictionnaire de l'ancien français: Le Moyen Âge*. Paris: Larousse.
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
- Hopper, Paul J. & Elizabeth C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Kellert, Olga. 2015. *Interrogative und Exklamative: Syntax und Semantik von multiplen wh-Elementen im Französischen und Italienischen*. Berlin: De Gruyter.
- Kellert, Olga. 2021. *Evaluative meanings of indefinites: A synchronic and diachronic analysis of Spanish cualquiera and related elements*. University of Göttingen. (Habilitation thesis).
- Kuryłowicz, Jerzy. 1965. The evolution of grammatical categories. *Diogenes* 51. 55–71.
- Lapesa, Rafael. 1975. El sustantivo sin actualizador en español: La ausencia del determinante en la lengua española. In Bosque Ignacio (ed.), *Presencia y ausencia de determinante en la lengua española*, 121–140. Madrid: Visor.
- Lehmann, Christian. 1982. *Thoughts on grammaticalization*. Munich: Lincom Europa.
- Lehmann, Christian. 1985. Grammaticalization synchronic variation and diachronic change. *Lingua e Stile* 20. 303–318.
- Menéndez Pidal, Ramón. 1928. *Manual de gramática histórica española*. Madrid: Espasa-Calpe.
- Meyer-Lübke, Wilhelm. 1899. *Grammatik der romanischen Sprachen*, vol. 3: Romanische Syntax. Leipzig: Reisland.
- Palomo, José R. 1934. The relative combined with *querer* in Old Spanish. *Hispanic Review* 2(1). 51–64.
- Pato, Enrique. 2012. Qual manera quier: La «interposición» en los indefinidos compuestos del español medieval. *Revista de filología española* XCII(2). 273–310.
- Rivero, María Luisa. 2011. Cualquiera posnominal: Un desconocido cualquiera. *Cuadernos de la Asociación de Linguística y Filología de la América Latina* 3. 60–80.
- Rivero, María Luisa. 1988. La sintaxis de qual quiere y sus variantes en el español antiguo. *Nueva Revista de Filología Hispánica* 36(1). 47–73.

- Stark, Elisabeth. 2006. *Indefinitheit und Textkohärenz: Entstehung und semantische Strukturierung indefiniter Nominaldetermination im Altitalienischen*. Tübingen: Max Niemeyer Verlag.
- Tobler, Adolf & Erhard Lommatzsch. 1925–2002. *Altfranzösisches Wörterbuch*. Peter Blumenthal & Achim Stein (eds.). Stuttgart: Franz Steiner Verlag.
- von Heusinger, Klaus. 2011. Specificity. In Klaus von Heusinger, Claudia Meinborn & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 2, 1024–1057. Berlin: de Gruyter.
- Zamparelli, Roberto. 2000. *Layers in the determiner phrase*. University of Rochester. (Doctoral dissertation).

Chapter 6

Indefinites and quantifiers in Old Sardinian: A corpus-based study

Guido Mensching

Georg-August-Universität Göttingen

The Sardinian systems of indefinites and quantifiers are interesting, among other reasons, because of the great proportion of loans from (Old) Italian, Catalan and Spanish – besides the preservation of a small number of Latin “archaisms” – as well as because of some interesting word order and agreement-related phenomena. As far as diachrony is concerned, a systematic analysis of Old Sardinian indefinites and quantifiers has never been undertaken. This chapter is based on a study performed by using the new corpus ATLiSOr, which has been available since 2017, and presents some first results on Old Sardinian negative indefinites and universal quantifiers. In particular, it turns out that the distribution and the frequency of some indefinites and quantifiers provide new insights into the issue of whether these elements are loans or whether they are inherited from Latin. The chapter also contains some first insights into the syntax of the items at issue and particularly examines the agreement behavior of *tot(t)u* ‘all’, a quantifier that is mostly invariable in Modern Sardinian. The study shows that agreement of adnominal *tot(t)u* was still optional in Old Sardinian and points out an interesting exception, namely that agreement was obligatory when *tot(t)u* was followed by a numeral.



1 Introduction

The systems of Modern Sardinian indefinites and quantifiers¹ are quite well known, mostly thanks to Jones's (1993) *Sardinian Syntax*; for Modern Sardinian indefinites, also see Mensching (2005). In contrast, very little research has been done on Old Sardinian, apart from Meyer-Lübke (1902), who dedicated one paragraph to indefinites (1902: 40–41), Wagner's (1938) *Flessione nominale del sardo antico e moderno*, where indefinites are dealt with on only four pages (1938: 128–132, §§40–46), and a section of Blasco Ferrer's (2003) analysis of the texts included in his anthology of Old Sardinian documents (2003: 207–208, §39: “Quantificatori”).² These contributions mostly bear on the inventory of forms and their origin, including the issue of borrowing, and do not say much on syntax.

Within the study of Romance indefinites and quantifiers, Sardinian is particularly interesting for at least two reasons: firstly, this language presents a remarkable number of indefinites and quantifiers that are loanwords from superstratum languages (Italian, Spanish, and Catalan), secondly, because some quantifiers show striking positional and agreement properties.

A thorough analysis of Old Sardinian has become possible only recently, after the online publication of ATLiSOr (*Corpus ATLiSOr: Archivio Testuale della Lingua Sarda delle Origini*) by Giovanni Lupinu in 2017. The aim of this chapter is to provide some first, mostly descriptive, results of a corpus analysis of Old Sardinian indefinites and quantifiers. More precisely, this study aims to (i.) assess the state of the art with respect to the origin of these items, and, in particular, the issue of whether they are loans or autochthonous elements, and (ii.) to thoroughly describe the syntax of some of these items for the first time.

The chapter is organized as follows: §2 contains some information on Sardinian (§2.1) and some notes on the state of research on Sardinian indefinites and quantifiers (§2.2). §3 is concerned with the negative indefinites *nemo(s)* ‘nobody’ and *perunu, niunu/neunu, nixunu/nexunu* ‘nobody/no (X)’. §4 focusses on the universal quantifiers *cada* and *omnia/omni/ogni* ‘every/each’ as well as *omnes* and *tot(t)u* ‘all’. The element *tot(t)u*, which is used to express the meaning of ‘all’ in Modern Sardinian, is widely documented in the medieval Sardinian texts. Today, it presents the striking property of being mostly uninflected for gender and

¹Note that, in traditional grammatical descriptions, quantifiers are subsumed under indefinites, but indefinites and quantifiers (and especially universal quantifiers) are usually kept apart in most modern linguistic frameworks, although both groups may share some properties and are sometimes diachronically derived from each other; see Haspelmath (1997: 11–13) for discussion. For generative frameworks, see Heim (1982), Beghelli & Stowell (1997), and Szabolcsi (1997). I would like to thank an anonymous reviewer for providing these references.

²In these works, quantifiers are treated together with indefinites; see footnote 1.

number, a trait that has never been the focus of any study of Old Sardinian. §5 therefore contains a study of the agreement properties of Old Sardinian *tot(t)u*. The results of this study will show that agreement of *tot(t)u* was still optional in Old Sardinian, except when it combines with numerals. This property will be discussed at the end of §5 from a typological perspective.

2 Sardinian

2.1 A brief history of Sardinian

Sardinian is the Romance language that developed after Sardinia came under Roman rule as a result of the First Punic War. It has survived until today in two main dialect groups (Campidanese in the south and Logudorese in the north, the latter including the linguistically conservative Central Sardinian or Nuorese). All varieties of Sardinian are in diglossia with the official language, Italian. The number of speakers is estimated to be approximately 1 million (cf. Moseley 2007).

Sardinia belonged to the Byzantine Empire from the seventh century, but was neglected by Byzantium due to attacks by the Saracens, which led to the development of independent political structures, the so-called Judicates of Cagliari, Torres, Arborea, and Gallura. The Saracen raids were finally stopped with the aid of Pisa and Genoa, who extended their power on the island, so that the Judicates (except Arborea) lost their autonomy (cf. Mensching & Remberger 2016: 270). The Italian dialects of these two city-states (i.e. Pisan, a variety of Tuscan, and Ligurian) constitute a first important superstratum for Sardinian, the second being Catalan, when the island was passed to the Crown of Aragon at the beginning of the fourteenth century. Catalan, in turn, was followed by (Castilian) Spanish after the unification of Castile and Aragon in 1479. In 1718, Sardinia was given to the House of Savoy and thus to Piedmont and to united Italy in 1861, leading to today's diglossic situation and a heavy linguistic influence through Italian (cf. Rindler-Schjerve 1987; for further references see Mensching & Remberger 2016: 270–217).

Old Sardinian is documented from around 1050 to around 1400. Strikingly, and in contrast to the medieval documentation of other Romance languages, the documentation is exclusively legal and administrative, i.e. there are no writings at all of other genres, such as literary texts (Wagner 1997: 80–83). According to Blasco Ferrer (1995: 250–251), these documents can be divided into three types: (i.) letters of the chancelleries of the four Judicates, mostly containing legacies and donations; (ii.) the *condaghes*, which are proceedings of transactions concerning property assets and housing stock of monasteries; (iii.) codifications of laws and municipal ordinances. The ATLiSO corpus covers the whole documentation of all three types of documents.

2.2 Old and Modern Sardinian indefinites and quantifiers

Table 1 shows some indefinites and quantifiers of Modern Sardinian together with their origin as assumed in Wagner's *Dizionario Etimologico Sardo* (DES) and additional literature.

Table 1: Some indefinites and quantifiers of Modern Sardinian

Inherited from Latin	
<i>tottu</i> 'all'	< Late Latin TÖTTUM (classical TÖTTUM, DES 2: 500–501)
<i>nudda</i> 'nothing'	< Lat. NULLA (DES 2: 175)
<i>nemos</i> 'nobody'	< Lat. NEMO (DES 2: 161)
<i>donza/dogna</i> 'every'	< Lat. ET OMNIA (DES 2: 188)
<i>meta/meda</i> 'much/many'	< Lat. META 'heap' (DES 2: 112)
Loans	
<i>calchi/carchi</i> 'some'	< Ital. <i>qualche</i> (DES 1: 269)
<i>cali(n)cunu</i> 'some(body)'	< Ital. <i>qualcheduno</i> (DES 1: 269)
<i>nessunu</i> 'no'	< Ital. <i>nessunu</i> (DES 2: 168)
<i>donzi/dogni</i> 'every/each'	< Ital. <i>ogni</i> (DES 2: 188)
<i>tzertu/certu</i> 'a certain'	< Ital. <i>certo</i> (DES 1: 447)
Uncertain	
<i>algunu/argunu</i> 'some(one)'	< Span. <i>alguno</i> (DES 1: 70–71), Cat. <i>algú/algún</i> or Lat. ALICUNUS?
<i>cada</i> 'every/each'	< Lat. CATA < Greek κατά (DES 1: 256, REW 1755) or via Span. <i>cada</i> ?
<i>perunu</i> 'nobody, no X'	< Old Italian <i>veruno</i> or Lat. PER ŪNUM? (DES 2: 251 vs. Blasco Ferrer 2003: 207)
<i>neunu/niunu</i>	< Old Italian <i>neuno</i> or Lat. NĚ(c) ŪNUM? (DES 2: 168–169 vs. Blasco Ferrer 2003: 207)

All these items already existed in Old Sardinian, except for *calchi/carchi* 'some'. Instead of *algunu* 'some(one)', Old Sardinian had *alicunu*, which Wagner considers as inherited from Lat. ALICUNUS, whereas he says that the more frequent Old Sardinian *alcunu* is probably already an Italianism (DES 1: 70–71). The preferred form today, *algunu/argunu*, comes from Spanish according to Wagner, who already documented *algunu* in the *Carta de Logu* of the Judicate of Arborea, a legal code issued in 1392, a date which would, however, indicate a Catalan rather than a Spanish origin. Note, however, that it is not excluded to consider both *alcunu*

and *algunu* as inherited from Latin with a syncope of *i* and preceding intervocalic sonorization in the second form. In the case of *cada*, Wagner (DES 1: 256) claims that intervocalic sonorization occurs in the Old Sardinian texts in which he found this form (CSMB and CV), so that he sees no reason to consider it a Hispanicism. In contrast, he supposes that *perunu* ‘no (X)’ stems from Ital. *verunu*, maybe with the influence of *per* (DES 2: 251), whereas Blasco Ferrer (2003: 207) wants to derive it from Lat. PER ŪNUM. Similarly, Blasco Ferrer suggests that *neunu* derives directly from Latin NĖ(c) ŪNUM, whereas Wagner believes it to be a loan from Italian. Today, *nessunu* or *nisciunu* are more widespread, which are clearly of Italian origin (DES 2: 168; Blasco Ferrer 2003: 207). In §3 and §4, I will assess most of these controversial cases on the basis of corpus data.

The quantifiers *meta/meda* ‘much/many’ and *tot(t)u* ‘all’ show an interesting syntactic behavior in Modern Sardinian. Whereas Sardinian indefinites usually occur preminally, the gender-invariable *meta/meda* can occur both pre- and postnominally. In prenominal position, it usually agrees in number by taking an -s in the plural, whereas for many speakers number agreement is lacking in postnominal position (Jones 1993: 36).³ For Old Sardinian, the corpus ATLiSOr has only three occurrences of *meta* (see (1) below), which had already been discovered by Wagner (DES 2: 112). As all are in the singular, nothing can be said about the agreement facts, but the two word orders are already attested, as can be seen from (1a) vs. (1b):⁴

- (1) a. (Cond. SPS 289, p. 252.2)
renovo custu co(n)dake [...], ki fuit de **te(m)p(us)** **meta**
‘I renew this *condaghe*, which has existed for much time’
- b. (Cond. SNT 1, p. 125.2)
pro meta servizu ki lis feki
‘for the great amount of service that I rendered to them’

³Examples:

- (i) metas/medas libros
(ii) libros meta(s)/meda(s)
‘many books’

⁴In this and the other examples, italics and round brackets come from the editors of the texts from which the examples were taken. For instance, in the examples in (1), the italics and round brackets stand for resolved abbreviations in the medieval manuscripts. I copied these markings as is from the ATLiSOr corpus. Please note that the corpus is composed of different editions with different standards (essentially round brackets vs. italics). Bold and underline are mine. I usually use bold to highlight the quantifiers and indefinites at issue and underline to highlight other properties mentioned in the explaining text.

- c. (StSS L. I-LXII, p. 30.24)
 çascatunu q(ui) aet cherre petha, **et paca (et) meta**
 ‘everybody who will want meat, either little or much’

Contrary to *meta*, the universal quantifier *tot(t)u* is widely documented in the medieval texts. In Modern Sardinian, adnominal *tot(t)u* is prenominal and followed by the definite article. It does not agree with the noun (cf. Jones 1993: 37), as the following examples show:

(2) Modern Sardinian (Logudorese)

- a. totu s’ abba
 all the water-F.SG
 b. totu s’ ozu
 all the oil-M.SG
 c. totu sos òmines
 all the-M.PL man-M.PL
 d. totu sas fèminas
 all the-F.PL woman-F.PL

However, as Jones (1993: 38) observes, in some dialects “a plural form *tottus* (invariable for gender) is used in some cases, in particular when this item occurs in isolation (see Farina 1973: 270),” see (3):⁵

(3) Modern Sardinian (Nuorese, Jones 1993: 38)

- Sun tuccàos tottus.
 are arrived-M.PL all-PL

Whereas in Campidanese, the ending *-us* corresponds to the regular masculine plural ending of nouns and adjectives with the singular in *-u*, this is not the case in Logudorese and Nuorese, where the paradigm is *-u* (sg.)/-os (pl.). This is particularly interesting against the background of the diachronic data, as regular masculine and feminine plural forms (*tot(t)os/tot(t)as*) are attested in Old Sardinian. I will return to this issue in §5.

⁵Instead, there is no agreement when *tot(t)u* precedes the participle:

(i) Sun tottu tuccàos. (Jones 1993: 38)

3 Negative indefinites

3.1 Distribution in the corpus and the issue of borrowing

This section focuses on the Old Sardinian negative indefinites *nemo(s)* ‘nobody’ and *neunu* and *nex(i)unu* ‘nobody/no (X)’ as well as *nullu* and *perunu* ‘no (X)’. Table 2 shows their distribution in the corpus.⁶

Table 2: Negative indefinites in Old Sardinian (absolute occurrences)

	<i>nemo(s)</i>	<i>nullu</i>	<i>perunu</i>	<i>neunu</i>	<i>nex(i)unu</i>
CVolg. AAC 1 (1066–1074)		1	4		
Priv. Log. (1080–1085)		1			
Carta arb. Gen. 2 (1112–1120)		1			
Montecass. 22 (1136)		1			
Montecass. 32-orig (1153)		2			
Montecass. 35 (1170)		2			
Montecass. 39 (1182–1183 ca.)	1	3			
Carta arb. (1184)			1		
Cond. Bar. II (1190)				2	
Carta don. (1211)			5		
CVolg. AAC 11–18 (1215–1217)			13	2	
Carta Ben. (1225)			4		
CVolg. AAC 19–21 (1225–1226)			14		
StSS (1316)				145	11
StCastel. (1334–1336?)				3	44
StCastel. 2 (1334–1336?)					1
CdLA (end of 14 th c.)		1 ⁷	11	2	24
Cond. SPS (end of 11 th –13 th c.)	1	1		8	
Cond. SNT (12 th –13 th c.)			1		
Cond. SMB (12 th –13 th c.)		6	10	1	1
Total	2	19	63	163	81

⁶In the left column, the texts that can be dated more or less exactly are arranged in chronological order, followed by another block of texts that extend over more than one century.

⁷The Italianism *nullu* ‘of no value’: “siat nullo et de neguna efficacia e valo(re)” (‘it shall be of no value and of no use’, CdLA LXXII 118, 9).

3.1.1 *Nemo(s)*, *nex(i)unu*, and *neunu*

Sardinian belongs to the few Romance languages that still preserve Lat. *NĒMO* (besides Romanian, Corsican, and some Tuscan dialects; see REW 5886 and Bertocchi et al. 2010: 81). Both the modern and the ancient form, according to the literature, is *nemos* (Mod. Camp.: *nemus*), where the -s is considered as analogical to other indefinites (*aliskis*, *uniskis*, cf. DES 2: 161, Wagner 1938: 131). Wagner only cites one example, corresponding to (4a) below, in which the item already appears with the analogical -s. The corpus ATLiSOr now shows one additional example without the -s, see (4b):

- (4) a. (Cond. SPS 68, p. 130.5)
 ki non bi aet bias **nemos**
 ‘that nobody is entitled [to possess him]’
 b. (Montecass. 39)
nemo no(n) ‘de-llis levet, ni(n) ambilla ni(n) pischi
 ‘nobody may remove [from the river] neither eels nor fish’

Strangely enough, these are the only examples in the whole corpus. The reason for this might be that ‘nobody’ is too general for juridical texts, which tend to be precise, using indications such as ‘no man’, ‘no woman’, etc. This is actually borne out in the texts, where such expressions are found in the majority of negative references to indefinite persons. Some examples are given in (5).

- (5) a. (CdLA XCVII, p. 136.2)
 <I>t<e>m hordinam(us) q(ui) **nexuna p(er)soni** de su regnu n(ost)ru
 d’Arborê no(n) usit nen deppiat deseredari sos figios [...]
 ‘Likewise, we order that no person of our Kingdom of Arborea must disinherit the sons [...]’
 b. (StSS L. I-XLIX, p. 25.35)
 Vivende su maritu, **neuna muçere** senza paraula dessu maritu suo
 pothat nen deppiat facher alcunu c(on)tractu
 ‘When the husband is alive, no woman can or may make any contract’
 c. (StSS L. I-LXXII, p. 33.8)
Neunu barberi radat sas d(omi)nicas nen i(n) festas solle(m)pnes
 ‘No barber may shave on Sundays nor on solemn holidays’
 d. (Carta don. 85, col. 1.29)
 Et icustu beni [...], non apat balia nin po[te]stadi **p(er)unu luigi (et)**
nin p(er)una p(er)soni, ki ad be(n)ni pust mei, a isfairi-llu
 ‘And (with respect to) this donation, no judge or no other person that will come after me can revoke it’

For the elements *nullu*, *neunu*, *nex(i)unu*⁸ and *perunu*, 164 occurrences in the corpus show the constellation with a noun marked [+human]. In other words, although *nemos* has survived until today, it was probably avoided in legal and administrative texts.⁹ However, this cannot be the only reason for the scarcity of *nemo(s)*, as pronominal *neunu* and *nexiunu*, both synonyms of *nemo(s)*, occur more frequently. Some examples are given in (6):

- (6) a. (StSS L. I-LXXVI, p. 34.43)
Et i(n) una hora **neunu** pothat occhier plus de unu a(n)i(m)ale grussu
'And in one hour, no one may kill more than one big animal'
- b. (StSS L. I-CXXXVI, p. 51.28)
Qui **neunu** c(om)poret casu ov(er) lana si no(n) i(n) sa platha.
'That no one should buy cheese or wool outside the market place.'
- c. (StCastel. CCXXI, p. 49.34)
Et qui **nexiunu** non poçat vendere assos predittos venditores
'And that no one can sell to the above-mentioned sellers'

Nevertheless, the pronominal use of *neunu* as in (6) (33 occurrences) is exclusively documented in the *Statuti Sassaresi* of 1316, a text that is known for its Italianizing tendencies. Pronominal *neunu* can therefore definitely be classified as an Italianism. Similarly, *nexiunu*, which has been clearly identified as an Italianism in the literature¹⁰ (cf. §2.2), in pronominal use, is found almost exclusively in the *Statuti di Castelsardo* (1334–1336?) (27 occurrences), with one additional occurrence of the variant *nixunu* in the *Carta de Logu* of Arborea (end of the 14th century). These two texts are also known for showing Italianisms. The items *perunu* and *nullu* do not occur as pronouns.

Neunu (and its variant *niunu*) and *nex(i)unu/nix(i)unu* were used also as noun-modifying (adnominal) negative indefinites. Most occurrences of adnominal *neunu* (119 out of 130) are found in the *Statuti Sassaresi* (1316), but unlike pronominal

⁸Other variants are *nessiunu* and *nensiunu*.

⁹The same holds for 'nothing', which has no expression in Old Sardinian texts.

¹⁰As for *nex(i)unu*, *nixunu*, *nisciunu* (where *x(i)* and *sci* represent [ʃ]), Wagner (DES 2: 168) argues that they derive from the Old Italian form *nexun(o)*, quoting Monaci (1955). However, the texts of this chrestomathy in which this form (as well as *nixun(o)*) appears all correspond to Lombard and Venetian dialects. Instead, Blasco Ferrer (2003: 207) claims that *nexiunu/nisciunu* are loans from the Pisan dialect. But note that Italian forms such as *nesciuno/nisciuno* do not seem to be Tuscan, but are rather found in Southern Italy (cf. Rohlfs 1969: 215). What seems most plausible to me is that these Sardinian forms stem from Ligurian *nesciun/nisciun*. Recall from §2.1 the influence of Genoa and the Ligurian dialect on Sardinian. In contrast, the variants *nessiunu* and *nensiunu* can stem from Tuscan.

neunu, it is also sometimes found in some other texts, with the earliest two occurrences found in the *Condaghe di Barisone II* (1190). The Italian influence on Sardinian started to become particularly palpable starting from the 13th century (cf. Wagner 1997: 234–235), but earlier influences cannot be excluded, so it seems likely that both pronominal and adnominal *neunu* are to be considered as Italianisms. Even clearer is the case with *nex(i)unu/nix(i)unu*, which is almost exclusively found in texts from the 13th to the 14th century,¹¹ with only one example that is possibly dated earlier (*Condaghe di Santa Maria di Bonarcado*, 12th–13th c.). This occurrence (*no li tolliant donamentu nixun fatu insoru* ‘they shall not take away any donation made to them’; Cond. SMB 33, p. 41.15) is a clear Italianism, as can be seen from the apocopated form *nixun*, which does not conform to Old Sardinian grammar. In the late *Carta de Logu* of Arborea, the masculine form even regularly shows the Italian ending *-o* instead of *-u*.

3.1.2 *Perunu* and *nullu*

In contrast to *neunu* and *nex(i)unu*, the item *perunu* already appears in the 11th century (*Carta volgare dell’Archivio Arcivescovile di Cagliari* n. 1, 1066–1074). Although the sea republics of Pisa and Genoa, after their victory over the Saracens in 1016, started to acquire privileges on the island over the course of the 11th century (see, e.g. the document known as *Privilegio logudorese* from the 1080s, cf. Wagner 1997: 233–234), it is extremely improbable that an Italianism (in this case *veruno*) appears integrated into Sardinian at such an early date in a form that shows irregular sound shift (*peruno*, see §2.2). I therefore tend to agree with Blasco Ferrer (2003: 207) that *perunu* should rather be considered as inherited from Latin. Blasco Ferrer does not account for the alleged etymon *PER ŪNUM*. I suggest that the origin of the Sardinian indefinite is rather **PERŪNUM*, with *PER-* being the well-known Latin intensifying prefix used with adjectives and verbs (cf., among others, FEW 8, 213–214). Such a formation would thus be semantically equivalent to *VĒRĒ ŪNUM* (> It. *veruno*), “an emphatically reinforced form of the so-called pronominal adjective *ūnus*, which often has an indefinite function” (Ramat 1997: 2). Another detail that speaks against an interpretation of *perunu* as an Italianism is the fact mentioned above that *perunu* is not used as a pronoun, unlike Italian *veruno* (cf. Ramat 1997: 9). Actually, as Table 2 shows, *perunu* is documented rather constantly until the 13th century, with the Italianism *neunu* only sporadically popping up, starting from 1190 onwards. From the 13th century onwards, *neunu* competes with a second Italianism, *nexunu*, which prevails in some texts.

¹¹StSS (1316), StCastel. (1334–1336?), CdLA (end of 14th c.).

As for *nullu*, with only 19 occurrences, it appears to have been scarcely used. Wagner (1938: 132) considers it as an Italianism, but strikingly it mostly occurs in the earliest texts,¹² starting from the 11th century, so it is either an autochthonous element or a Latinism.

3.2 Syntax

The examples in (4) in §3.1.1, repeated here as (7a,b), show that *nemo(s)* undergoes negative concord¹³ both in postverbal and in preverbal position:

- (7) a. (Cond. SPS 68, p. 130.5)
 ki non bi aet bias **nemos**
 ‘that nobody is entitled [to possess him]’
 b. (Montecass. 39)
 nemo no(n) ’de-llis levet, ni(n) ambilla ni(n) pischi
 ‘nobody may remove [from the river] neither eels nor fish’

This suggests that Old Sardinian was a strict negative-concord language,¹⁴ unlike modern Sardinian, which shows the negation only when the negative indefinite is postverbal (cf. Jones 1993: 23). This can be confirmed with the items *nullu* and *perunu*, which are always accompanied by *non* ‘not’ or *nen/nin* ‘neither/nor’, even in the rare cases in which they occur preverbally (10.5%; *nullu*: 2 out of 19 cases; *perunu* 4 out of 38 cases¹⁵). The fact that these items occur mostly in postverbal position is not surprising, given Wolfe (2015) finding that Old Sardinian was fundamentally a V1-language (an insight to which I will return). Here are some examples for postverbal and preverbal *nullu* and *perunu* in (8a,b) and (8b,c) respectively (NEG and V underlined):

¹²With one exception in CdLA LXXII, 118.9, in which, however, the ending *-o* identifies the element as an independent Italianism. In addition, here, the whole construction *X siat nullo* ‘X be of no avail’, in which *nullo* is not adnominal (contrarily to the Old Sardinian use), actually calques an Italian model.

¹³The term “negative concord” refers to cases in which there is “a single interpretation of negation in the face of multiple apparent *negative* exponents” (Giannakidou 2020: 458).

¹⁴In “strict negative-concord languages”, a negative marker is obligatory with negative indefinites, independently of their position. Cf. Giannakidou (2000, 2006), among many others. Note that Latin did not have negative concord, but was rather a “double negation” language, in which two negative items yielded a positive reading, cf. Gianollo (2016).

¹⁵I counted only the occurrences of *perunu* with a negative meaning. For positive *perunu* see below.

- (8) a. (Cond. SMB 17, p. 20.17)
 Et non appat ausu **nullum hominem** a ttollerendellos aligando de
 servitiu de *sancta* Maria.
 ‘And no man dare (lit. not-have.3SG-SUBJV daring no man) to take
 them off the service of Saint Mary.’
- b. (Carta Ben. 93, 1.17–18)
 [Et] no-ndi levit **pegus perunu** pro terra maina (et) ni **atera causa**
p(er)una
 ‘And he may not take away any cattle (lit. NEG=from.it
 take.away.3SG-SUBJV cattle any) for clayey soil nor for any other
 thing’¹⁶
- c. (Montecass. 39)
 (et) pischi **nullu ho(m)i(n)e** mortale no(n) ‘de-llis levet
 ‘and no mortal man may remove fish from them’
- d. (CdLA V, p. 60.6)
 siat i(n)furchadu qui ‘(n)di mo(r)giat e **p(ro) dinari p(er)u(n)u** no(n)
canpit
 ‘(he) shall be hanged so that he dies, and he cannot live for any
 money’

Unlike *nemo(s)* and *nullu*, *perunu* has a positive meaning (‘any’) in irrealis contexts in 9 cases out of 47 (19.1%), both in preverbal (5 occ.) and in postverbal (4 occ.) position, so *perunu* should rather be classified as a negative polarity item (NPI). Here are two examples:

- (9) a. (CVolg. AAC 19, p. 313.18)
 Et si **p(er)unu tempus** illoi bolint torrari hominis ad istari in cussa
 billa, [...]

 ‘And if (at) any time persons want to return here to live in this village,
 [...]’
- b. (CVolg. AAC 20, p. 315.14)
 et issu et totus sus piscobus [...] bollant pasquiri cu(m) peguliu issoru,
 bollant arari, [...] , ho piscari, ho fayri **peruna atera causa**
 ‘and he and all the bishops [...] may pasture their cattle, may plow, or
 fish, or do any other thing’

¹⁶The sense of this sentence is not clear. For *terra maina* ‘black/clayey soil’, see Blasco Ferrer (2003: 96, 237).

This is a clear parallel to Italian, where *veruno* developed from an NPI to a negative indefinite. Nevertheless, this fact need not be attributed to Italian influence, given the similar origin of *veruno* and *perunu* (see §3.1.2), with neither of the two items containing a negative element. As Ramat (1997: 4) points out (citing Haspelmath 1997: 222), in several languages, indefinites made up of an emphasizing or focalizing element and an item meaning ‘one’ have developed an exclusively negative meaning over time. This actually happened to both Sard. *perunu* and It. *veruno* in the modern stages of Sardinian and Italian, respectively.

Let us now look at *neunu* and *nex(i)unu*. Like Sardinian *perunu* and the Old Italian items *neuno* and *nesciuno* (see Franco & Poletto 2016, Franco et al. 2016), Sardinian *neunu* and *nex(i)unu* sometimes show an NPI-like behavior. This occurs when they are used in postverbal position without negative concord,¹⁷ in which case they do not have a negative meaning, thus corresponding to English ‘any(body)’. However, this is found only eight times with *neunu*+N, almost always in sentences that express a condition and that contain the verb *kertare* ‘to bring a lawsuit’ as in (10a). These eight occurrences are found in three texts, namely the *Condaghe di Barisone* (1190), the *Condaghe di San Pietro di Silki* (end of 11th–13th c.) and the *Condaghe di Santa Maria di Bonarcado* (12th–13th c.). NPI-like *nexunu* appears only in one conditional clause but with a different verb, see (10b), from the *Statuti Sassaresi*.

- (10) a. (StSS L. I-CXVI, p. 48.34)
 cambiando cun boluntate de pare e bocando·nde de si ’nde kertavat
neunu homine mortale o isse, d’ispiaremila a ssanta Maria
 ‘interchanging [these properties] in joined commitment and with the
 condition that, should any mortal or he (himself) bring a lawsuit
 concerning this matter, it will be exempt from claims in favour of St.
 Mary’
- b. (StSS L. II-59, p. 84.44)
 E si li ma(n)chat **bestia nexuna**, [...], si paguet dae sos benes de su
 dictu comunargiu minore
 ‘And if any animal is missing, [...], it shall be paid from the
 belongings of the minor herdsman at issue’

Now let us turn to negation of *neunu* and *nexuno*. Examples like (5a) and (6c), in which *nexunu* occurs in a preverbal position, seem to confirm the status of Old Sardinian as a strict negative concord language, as does (11) for *neunu*:

¹⁷In Old Italian, such occurrences of negative indefinites with an NPI-reading occur in a broader set of contexts (in addition to conditions, see also hypothetical free relative clauses and questions).

- (11) (StSS L. II, p. 63.6)

Qui **neunu** **corssu** no(n) pothat aver officiu i(n) sa citadi de Sass(ar)i.
 ‘That no Corsican may hold office in the city of Sassari.’

However, in strong contrast to *perunu* and *nullu*, these items often lack the negation in preverbal position, even when they are clearly negative, like in (6b) (repeated as 12a) and (12b):

- (12) a. (StSS L. I-CXXVI, p. 51.28, repeated from (4b))

Qui **neunu** c(om)poret casu ov(er) lana si no(n) i(n) sa platha.
 ‘That no one should buy cheese or wool outside the market place.’

- b. (StCastel. CCXXVI Rubr., p. 51.1)

Qui **nexiunu** vendat vinu a barile.
 ‘That nobody may sell wine by barrel.’

Actually, the presence of the negating element as in (11) is quite rare with *neunu*, in contrast to adnominal *nexunu*, where 20 out of 36 relevant cases show the negation. In any case, the optionality of negation with preverbal n-words¹⁸ is a property that is typically found in Old Italian (see Franco & Poletto 2016; Franco et al. 2016). It therefore seems that these items are not only loans from Italian, as the quantitative data presented in §3.1.1 suggest, but their syntax, too, is a calque from the corresponding Old Italian structures.

Strikingly, pronominal *neunu* and *nexunu* do not occur at all in the postverbal position. When these negative elements are used as modifiers, they are attested postverbally, but with an extremely low frequency. For *neunu*, only the following four (out of 163 occ.) could be identified:

- (13) a. (Cond. SPS 410, p. 330.19)

(et) simile no(n) possa(n)t laorare miglaresos, butones né **neunu** **at(er)u** lauru qui siat minus dessa tocha [...]
 ‘and similarly they cannot produce (either) *miglaresos*,¹⁹ (or) buttons or (lit.: nor) other products that are inferior to the alloy [...]

- b. (Cond. SPS 410, p. 330.19)

plachit a donnu Saltaro de Cherchi accordaresende cun sos donnos **kene kertu neunu**
 ‘it pleased Donnu Salataro de Cherchi to come to an agreement with the gentlemen without any legal dispute’

¹⁸I follow the terminology of Franco & Poletto (2016: 1), according to which n-words are “words morphologically starting with the negative morpheme *n-*”

¹⁹A type of silver work. Cf. Tola (1850: 128).

- c. (StSS L. I-CXXV, p. 51.26)
ma cussu sacrame(n)tu siat tentu de facher, sensa neuna
adpellatio(n)e.
'but he shall be obliged to do this sacrament without any appeal.'
- d. (StSS L. II-VI, p. 66.29)
et i(n) cussa q(ue)stio(n)e se p(ro)cedat, sensa neunu atteru term(en)
'and one shall proceed in his issue without any other delay'

In each of these cases, the postverbal negative indefinite is licensed via negative concord, as is expected both in Old Italian and in Old Sardinian, either by *non/ne* or by *kene/sensa* 'without'. For *nexunu*, out of a total of 68 occurrences, there are only 11 occurrences in which this item appears postverbally: one with a postverbal direct object similar to (13a) ("no li tolliant donamentu nixun fatu insoru" 'they shall not take away any donation made to them', Cond. SMB 33, p. 41.15, already cited in §3.1.1), and two cases of prepositional phrases with an adverbial function (*i[n] nessiunu modu/per modu nixunu* 'in no way', StSS L. I-CLII, p. 59.38 and CdLA CXLIX, p. 184.4). These three occurrences are licensed by *non*. In addition, there are three occurrences with *sensa* 'without' similar to (13c–d). Finally, there are eight cases of the following type, always with an expression meaning 'of no value', which show the lack of negative concord:

- (14) (StSS L. I-CXVI, p. 48.34)
Et si c(on)tra aet ess(er) factu, cussa accusa siat de nessiunu valore
'And should this be disregarded, the respective accusation will be of no value'

This structure mirrors the behavior of Old Italian *n*-words in contexts with the meaning 'no/little value' (cf. Franco et al. 2016) and can also be considered as an influence of Italian.

The extremely low frequency of postverbal occurrences of the elements at issue diverges from Old Italian, where postverbal negative indefinites are frequently found. This is even more puzzling against the background of the behavior of *nullu* and *perunu*, and, generally, of Wolfe's (2015: 20–21) findings on Old Sardinian syntax:²⁰ According to his study, Old Sardinian shows V1 word-order in around 73 percent of matrix clauses and all of the embedded clauses. Matrix clauses (but not embedded clauses) also appear with V2 (25%) and – very

²⁰In this study, Wolfe analyzes extracts from the *Condaghe di San Nicola di Trullas* and the *Condaghe di Santa Maria di Bonarcado*.

marginally – with V3 (1.7%) and V4 (0.5%) order. Now, interestingly, in the sentences or clauses that contain the two indefinites at issue, V1 order only appears in the rare NPI cases mentioned above and in some of the cases with postverbal (negative) *neunu* and *nixunu*.

Table 3: Word order in sentences containing *neunu* and *nexunu*

	<i>neunu</i>	<i>nexunu</i>
V1	10 (5.9%)	10 (13.5%)
V2	114 (67.5%)	57 (77.0%)
V3	32 (18.9%)	4 (5.4%)
V4	13 (7.7%)	1 (1.4%)
V5	0 (0%)	1 (1.4%)
V6	0 (0%)	1 (1.4%)
Total	169	74

Examples for V2 order can be seen in (5a,c), (6b,c), (11), (12a,b); for V3 and V4 order, see (6a) and (5b), respectively. (15) is an example with embedded V6 order. This is the only example with such order, which can be considered as ‘extreme’ in that sense:

- (15) (StSS L. II-XXXXI, 80.36)
ordinait qui [₁ dae como inantis] [₂ su pot(estate) q(ui) e(st) e pro
temp(us) at ess(er)], [₃ a req(ue)sta **de nexiunu creditor(e)**] [₄ **nexiunu
corp(us) mortu**], [₅ **p(er) nexiunu deppidu** de qualu(n)cha qua(n)titade
siat obligadu], no(n) si poça(n)t ne(n) deppia(n)t, i(n) sa dicta citade [...]
staxiri ne(n) inpedire
‘he ordered that, [₁ from now on], [₂ the potestate who is or will be in
charge] [₃ on request of any creditor] cannot and must not either
confiscate or block [₄ any dead body] [₅ for any debt of whatever
quantity might be owed].

In the data examined here, there is thus no matrix/subordinate clauses asymmetry with respect to V>1 word-order, unlike what Wolfe found in his corpus. These variations from V1-syntax are almost exclusively restricted to the *Statuti Ssassaresi*, the *Statuti di Castelsardo* and the *Carta de Logu d’Arborea*, three texts where Italianizing tendencies are expected, which, in this case, affect a core syntactic parameter. How about the puzzling fact that the relevant items almost

never occur postverbally when they have a negative meaning? A future study might investigate whether this is due to the ambiguous status of these items in Old Italian (i.e. they could either undergo negative concord or function as NPIs). More particularly, it might be that in the postverbal position, these items were borrowed almost exclusively in their positive meaning.

4 Universal quantifiers

4.1 *Cada*

The item *cada*, which is widespread in Modern Sardinian in the meaning ‘every/each’, ultimately derives from Greek κατά in its distributive meaning. The FEW (2:482) particularly mentions its use with time indications and numbers: κατά μήνα ‘every/each month’, κατ’ ἐνιαυτόν ‘every year’/‘yearly’, καθ’ ἓν ‘one by one’, κατὰ τρεῖς ‘three each’ (also cf. Rohlfs 1969: 220, Hofmann 1972: 254–255). The item existed as a loan in Latin, starting from the 3rd century AD, with the same distributive meaning: *cata mane mane* ‘morning by morning’, *plica unum cata unum petalum* ‘fold the gold threads one by/after one’ (FEW, loc. cit.). This distributive use can still be found in some Romance varieties (e.g. in Romanian, cf. REW 1755). In contrast, in Ibero-Romance and in Sardinian this item developed the meaning ‘every/each’. In other varieties, this meaning is only found in combinations with the word for ‘one’, like Old Northern Italian *cad(a)uno*, Old Tuscan *catuno*, with variants such the Old Pisan *cateunu* (‘each one/everyone’, cf. Rohlfs 1969: 220–221).

Wagner (DES 1: 256) claims that both Modern and Old Sardinian *cada* is an inherited form from Latin. But note that it is expected that the item at issue be pronounced */kata/ in most of the modern central (Nuorese) dialects (which do not show sonorization of intervocalic Latin voiceless plosives).²¹ This is not borne out, yet: the sonorized form /kada/ is found everywhere. Therefore, Wagner suggests that the pronunciation in the modern central varieties may have been influenced by Spanish. Our discussion below suggests another picture, according to which Old Sardinian *cada* may directly stem from Latin, whereas its Modern Sardinian equivalent is most probably a loan from Catalan or Spanish.

As for the medieval documentation, Wagner (DES 1: 256) says that *cada* figures various times in the *Carte volgari dell’Archivio Arcivescovile di Cagliari* and the *Condaghe di Santa Maria di Bonarcado*, quoting the examples *cada VIII sollos*, lit.

²¹The change from -t- to -d- in the Old Sardinian texts in which *cada* occurs is not a problem, as intervocalic sonorization is a regular phenomenon in these texts (cf. DES 1: 256).

‘every/each 8 silver coins’ (but see below), *ankilla de cadadie* ‘maidservant in full time/full possession’, and *serbus de cadadie* ‘serfs in full time/full possession’.

Looking at the corpus, we see that *cada* as an independent word only occurs once:

(16) (Cond. SMB, 67, p. 55.3)

Ego Cipari de Lacon avia saltu cun sanctu Augustinu et cun donnigella Maria, cia mea, **cada** .VII.²² sollos.

‘I, Cipari de Lacon, possessed some (wood)land together with (the monastery of) Sanctu Augustinu and the princess Maria, my aunt, (worth) 7 soldos **each**.’

This is a clear distributive use, in which *cada* does not modify a noun but is rather used adverbially in the sense of ‘in each case’, ‘for each’, ‘respectively’, and resembles rather the Greek usage as in *κατὰ τρεῖς* mentioned above. As Grandgent (1907: 37) assumes, “*cata* was probably introduced, along the Mediterranean, by Greek merchants, in such [Latin] phrases as *cata unum* = *καθ’ ἕνα*, *cata tres* = *κατὰ τρεῖς*.” In any case, the isolated occurrence of *cada* in (16) does, by all probability, not attest to the existence of a universal quantifier *cada* in Old Sardinian, but is rather a reflex of Latin or even Byzantine Greek formulaic bookkeeping language. As for Greek, recall that Sardinia originally belonged to the Byzantine Empire (cf. §2.1). The Judicates took up Byzantine administrative structures, and, as Wagner (1997: 165–174) demonstrates, the Old Sardinian chancellery language has multiple influences of Byzantine Greek.²³ Thus, this isolated occurrence of *cada* does not correspond to the modern adnominal use (*cada* X ‘every/each X’).

²²The modern edition used in ATLiSO does say “VII”, whereas Wagner (see above) writes “VIII”, probably following an older edition.

²³Also note that another element, *cana*, is slightly more frequent (4 occ.) and is used in exactly the same way as *cada* in (16):

(i) (Cond. SMB, 100, p. 71.25)

Partirus fijos de Justa de Scala et de Eizu de Esule: et levarus fijos de cussos **cana** .II.

‘We divided the sons of Justa de Scala and of Eizu de Esule: and we took from them two each.’

Meyer-Lübke (1902: 70) proposed that this item is a blend of Lat. *cata* (see above) and *ana*, a late Latin loan from Greek *ἀνά* with the same distributive function as *κατά*; for Latin, see Hofmann (1972: 254) and, particularly, the example *ana duas tunicas* ‘two tunics each/apiece’. Given the slightly better documentation of *cana*, it is therefore not excluded that *cada* in (16) is a scribal error for *cana*.

All other occurrences of *cada* occur inside of what seems to be a compound word, *cadadie/cadadia*, and are found exclusively in some documents of the *Carte volgari dell'Archivio Arcivescovile di Cagliari* (dated 1215 and 1217), and always in the fixed expressions mentioned by Wagner: *ankilla de cadadie* (3 occ.), *serbu(s) de cadadie* (5 occ.) and *serbus de cadadia* (1 occ.).²⁴ The use of *cada* in a compound does not prove the existence of this item as a quantifier in the Old Sardinian texts.²⁵

It thus seems that a universal quantifier *cada* 'every/each' is not attested in Old Sardinian. Instead, several derivations of Lat. OMNIS were used, which will be the subject of the next subsection. The modern Sardinian quantifier *cada* is thus quite clearly a later loan from Catalan or Spanish.²⁶

4.2 Forms related to Lat. OMNIS

In the medieval Sardinian texts, several forms related to Latin OMNIS can be found, which are shown in Table 4.

²⁴The form of the item in the latter expression (*cadadia*) is strange, as the word for day is *die* in the whole documentation of Old Sardinian. It looks like a Hispanicism, which would be unexpected for this text, as the Catalan and Spanish influence did not take place before the 14th century (Sardinia belonged to Aragon from 1326 onwards). But note that this form seems to stem from a 15th c. copy. CVolg. AAC 11–21, whose edition was used for the corpus ATLiSOr, adds a footnote explaining that the parchment itself writes *cadaia* (CVolg. AAC: 306 11–21). This form is probably corrupt and hence not conclusive.

²⁵The compound itself is still enigmatic, an issue that cannot be resolved here. The development of *-t-* > *-d-* would indicate that the compound is an older lexicalization of a Latin **cata diem* that underwent this sound change. Alternatively, we could assume that the *-d-* in *cada* is due to a long-distance assimilation to the *-d-* in *die*. Finally, I would not exclude that *cadadie* is a loan-blend of Greek καθ' ἡμερᾶν/καθημέραν or καθημερινός 'daily' (Sophocles 1900: 612), in which the *-d-* could stand for the interdental fricative *-θ-*. Note that Solmi (CVolg. AAC: 319), in his lexical notes, says that CVolg. AAC 16, 307.5 has *de catadie*, but the text says *de cadadie*, and there is no note specifying a variant.

²⁶A whole series of loans stemming from or related to Lat. *cata* appears in the 14th century (*Statuti Sassaresi*, *Statuti di Castelsardo* and *Carta de Logu d'Arborea*). The corpus shows *casc(h)unu/-a*, *ciascunu/-a* and *casc(h)adunu/-a*, *ciasc(h)adunu*, *ciascatunu*, *-a*, *çascatunu*, *-a*, *ciascu->unu/çascadunu*, *-a* 'every/each', more rarely 'everybody', all from O. Italian (see Rohlfs 1969: 220–221 for these and/or similar forms). The Sardinian loans sometimes appear with the original It. ending *-o* instead of *-u*. In contrast, *cadiscuno*, *-a* is restricted to the *Carta de Logu d'Arborea* (end of the 14th century), where it occurs three times. It appears to be a loan from Cat. (*cadascú/cadescú*) or O. Sp. *cadascuno* (cf., among others, Malkiel 1948: 396) or *cade-scuno*, which is rather O. Arag. (documented in a text of 1385–1396 edited by Cacho Blecua 2003).

Table 4: Occurrences of forms related to Latin OMNIS

	<i>omnia/ onna</i>	<i>do(n)nja/ donnia</i>	<i>(d)ogna</i>	<i>omni/ donnj</i>	<i>(d)ogni</i>	<i>omnes/ onnes</i>
CVolg. AAC 1 (1066–1074)	2					3
Priv. Log. (1080–1085)						3
Montecass. 9 (1082–1112)						3
Carta arb. Gen. 2 (1112–20)	2			1		
Montecass. 10 (1113)	1					3
Montecass. 5 (1120)						4
Montecass. 12 (1120?)						3
Montecass. 16 (ca. 1120)	1					3
Montecass. 20 (1134?)						2
Montecass. 22 (1136)				1		3
Montecass. 32-orig (1153)						3
Carta gall. (1173)	3					
Montecass. 39 (1182–83 ca.?)	3					1
Carta arb. (1184)						1
Cond. Bar. II (1190)	5	1				
Carta don. (1211)						
Tratt. Pace (1206)						1
CVolg. AAC 11–18 (1215–17)	36					
Carta Ben. (1225)						
CVolg. AAC 19–21 (1225–26)						
StSS (1316)	26		10	3	15	
StCastel. (1334–1336?)			2	1	1	
CdLA (end of 14 th c.)	1	20	3	1		
Cond. SPS (end of 11 th –13 th c.)	31	3		1		
Cond. SNT (12 th –13 th c.)	15					
Cond. SMB (12 th –13 th c.)	100	6		3		9
Total	226	30	15	11	16	44

The most frequent form is *omnia* and its variant *onna* ‘all/every/each’ (< Lat. OMNIA), which Wagner (1938: 129–130) considers an Italianism, given that it preserves the Latin *-i-* (cf. Wagner 1938: 130).²⁷ By this he seems to mean that *omnia/onna* are Latinizing spelling variants of O. It. *ogna*,²⁸ and he observes that the Italian spelling *ogna* is often found in Old Sardinian texts. As Table 4 shows, this is, by far, not the case, given the only 15 occurrences of *ogna* documented (vs. 226 occurrences of *omnia/onna*). In addition, while *ogna* is frequently found in Old Pisan texts (165 occ. in the OVI corpus), there is no trace of the spellings *omnia*²⁹ and *onna*. Therefore, either *omnia* is a Latinism with a popular adaptation *onna*, or *onna* represents an inherited form, of which *omnia* was a Latinizing spelling. The latter seems more probable, as (like O. It. *ogna*) the originally neuter plural item is used here as a singular quantifier with the meaning ‘every/each’ preceding both masculine and feminine forms, as is shown in (17).³⁰

- (17) a. (CVolg. AAC 1, p. 43.2.2)
 (et) fazzant **o(mn)ia** serbiciu (m.)
 ‘and they shall provide every service-M.SG’

²⁷I think that this view is not conclusive. As OMNIA must have yielded a Vulgar Latin *[ɔn:ja], we have to look at other Vulgar Latin words with [-nj-]. An example is V. Lat. *[βinjas] (Lat. VINEAS), which actually shows the *-i-* (probably representing [j]) in O. Sard. *vinias* (besides Latinizing *vineas*) and the Italianizing spelling *vi(n)gnas*, which is particularly frequent in the *Statuti Sassaresi* and the *Statuti di Castelsardo*. The former shows some isolated cases of *vingias*, where *-gi-* represents a palatal affricate, a later Sardinian development. It is probable that *-j-* in *vinjas* in the *Carta de Logu d’Arborea* stands for the same sound. The fact that this development is usually not reflected in the results of OMNIA may be accidental or result from an impact of Latin spelling (but see *donja* in the *Carta de Logu d’Arborea*).

²⁸According to Rohlfs (1969: 219), *ogna* is mainly a Lombard, Venetian and Northern Tuscan form, thus being extremely rare in Old Florentine (only 8 occ. vs. 1042 occ. of *ogne* and 11065 occ. of *ogni*), whereas it is more frequent in Old Pisan (*ogna*: 165, *ogne*: 184, *ogni*: 4051), according to the OVI corpus.

²⁹All 37 occurrences found in Old Pisan texts in OVI appear exclusively in Latin quotations. As a Romance element, *omnia* is extremely rare in all medieval Italian dialects, although some isolated cases can be found in Old Lombard, Old Venetian, and Old Umbrian, all dialects that did not have any impact on Sardinian.

³⁰It occurs very rarely with a plural noun: *de om(n)ia maioraes suos de locu* ‘(of all members of the leading families of the place’, *Carta Arb. Gen.* 2 [1112–20], 104, 2.17), *et cun omnia libertatos suos* ‘(and with all its freed serfs’, *Cond. SMB* [12th–13th c.] 1, 9.11 and similarly in *Cond. SMB* 207, 131.22). In the Old Pisan texts, *ogna* also rarely occurs with a plural noun, but only for indicating time intervals (of the type *ogna sei mesi*). It is therefore probable that the expressions *om(n)ia sex meses* ‘every six months’, *StSS L. I-XLVII*, 25.21 and *om(n)ia duos me[s]es* ‘every two months’, *StSS L. I-XXVIII*, 13.33; similarly in *L. I-XCIX*, 40.39) are calques from O. It.

- b. (Carta gall., p. 177, col. 1.28)
 cu(n) **omnia** p(er)tine(n)thia issoro (f.)
 ‘with all their possession-F.SG’

Interpreting *omnia* as a singular, is, in fact, a Vulgar Latin innovation that can be traced back to at least the 2nd c. AD, as the following example from Norberg (1944) shows:³¹

- (18) Late Latin (CIL 1, 583, 73, cf. Norberg 1944: 55, quoted in Rohlfs 1969: 219, fn. 2)
omnia quod ex hac lege factum non erit
 ‘all that will not have been done following from this law’

Notably, a pronominal use in the sense of ‘all, everything’ is not found with O. Pisan *ogna*, in contrast to Old Sardinian, in which *omnia/onna* can be used as a pronoun, but only when it is restricted by a relative clause, i.e. exactly in syntactic contacts like (18). However, as a relative pronoun, this language used *cantu* (< Lat. QUANTUM, DES 1: 289), thus diverging from the Latin construction.³² This structure makes up around 55 percent of the occurrences in the corpus (124 of 226 occurrences). It is generally found in the shape [*omnia/onna cantu* ... V], as in (19a,b), with only two exceptions: In (19c), the relative pronoun is *ca* instead of *cantu*, and in (19d) the conjunction *et* ‘and’ is located between the quantifier and the relative pronoun.

- (19) a. (Cond. SMB 36, p. 47.10)
 Et confirmolli sa domo de sancta Barbara de Turre cum **omnia cantu**
 aet cun terras cum binias cum servos et ancillas
 ‘And I confirm to him the [possession of the] house of Saint Barbara
 de Turre with all that it has, [i.e.] with vineyards, with serfs and
 maids’
 b. (Carta gall., pag. 177, col. 1.28)
 Co(m)porai-li a Gavini de Vare, su p(re)viteru de Bosove, **o(nn)ia**
ca(n)tu vi avet i(n) balle de Bosove dave su molinu de Castra i(n) iosso
 ‘I bought from Gavini de Vare, the priest of Bosove, all that there is in
 the valley of Bosove from the mill of Castra downwards’

³¹The remarks by Rohlfs (1969: 219) suggest that O. It. *ogna* is due to the fact that the *-a* in *omnia* was interpreted as a feminine form, and hence *ogna* is used with feminine nouns only. This is not true for the Old Pisan texts, which show both genders according to the OVI database. However, Rohlfs may be right for other Italo-Romance varieties, e.g. Old Venetian, for which the OVI corpus shows the feminine, with very scarce exceptions.

³²The (written) Latin construction would be *omnia qua*, or, in any case, *omnia quanta*. Wagner (DES 2: 188) says that this structure might be a Latinism, but he is referring to *omnia* itself, which has preserved the Latin meaning of ‘all’ in the sense of ‘everything’.

- c. (Cond. SPS 139, 170.2)
 MAXIMILLA ABBATISSA dessu monasteriu de S(an)c(t)u Pet(ru) de
 Silki, ki ponio in ecustu condake pro **o(mn)ia ca** 'nke parai in sa domo
 'MAXIMILLA, ABBESS of the monastery of Saint Peter of Silki,
 whom I mention in this *condaghe* for all that I acquired in the house'
- d. (CVolg. AAC 20, 316.4)
 Et daulloy assu do(n)nu miu s(an)c(t)u Antiogu d'iscla de Sulchis
o(m)nia et cantu apu dessu saltu
 'And I give to my lord, Saint Antiogu of the island of Sulchis, all that I
 have of the (wood)land'

The fact that this construction is totally absent from Old Italian can be taken as proof that *omnia/onna* is not an Italianism. This argument is further corroborated by the data of the clear Italianism *(d)ogna*, which appears late (14th c.), and exclusively in the three texts that we have already identified as being prone to Italianisms, the *Statuti Sassaresi*, the *Statuti di Castelsardo*, and the *Carta de Logu d'Arborea*. Notably, **(d)ogna cantu* is not attested, which falls in place due to the absence of similar structures in Old Italian. The initial *d-* in *dogna*³³ (4 cases in the *Statuti Sassaresi*) is not restricted to the Italianism, but occurs earlier with the autochthonous element *omnia/onna*, starting from the end of the 12th century both as a pronoun followed by *cantu* as in (20a) and as a modifier, see (20b, c). In the late *Carta de Logu* it is almost generalized and appears in the spellings *donnja* and *donja*, like in (20c), where the *-j-* may represent a palatal affricate already indicating the modern development of Vulgar Latin /nj/ to /ndʒ/ reported in Wagner (1907: 58).

- (20) a. (Cond. Bar. II II, p. 63.22)
 Ego, iudike Barusone, conp(or)ai-li a Mariane de Varru su de Usone
do(n)nia ca(n)tu bi aviat in I(n)nobiu de vineas, et terras, et saltos,
 (et) corte, et ho(m)i(n)es.
 'I, the judge Barusone, bought from Mariane de Varru, the one from
 Usone, all that there was in Innobiu of vineyards, (wood)land,
 courtyards and people.'

³³Wagner (DES 2: 188) accounts for this *d-* as the result of a wrong segmentation of *(d)ed omnia*, which, according to him, is frequently found in the Old Sardinian texts. However, the corpus shows no occurrences of this string. But a similar hypothesis is possible for *et omnia/onna*, considering that the final *-t* of *et* could be sonorized before vowels (see the alternative spelling *ed* in this phonological contexts in the corpus).

- b. (Cond. SMB, 30, 38.2)
cun lassando parte a ffilios e a donnia frate suo
 ‘by leaving a part to the sons and each brother of theirs’
- c. (CdLA CV, p. 144.15)
*q(ui) ad bendere cu(n) att(e)ra misura si no de cusas qui naradas
 su(n)t paguit p(er) donja volta (sollos) VI*
 ‘who sells using a measure other than those which are listed shall pay
 for each time six soldos’

The form *omni* is rarely found, and partially appears in Latin or Latinizing formulae, such as in *om(n)i opera bona* (Carta Arb. Gen. 2, p. 104, 2.25), *de o(mn)i op(er)a* (Montecass. 22, p. 170, 1.10), *cessante omni iustu impedimentu* (‘every legal obstacle having ceased to exist’, StCastel. CLXV, p. 37.24).³⁴ Here, *omni* seems to be a fossilized Latin ablative. The same could be said for the pronominal use in *cum omni cantu at* (‘with all that he has’, Cond. SMB 36, p. 46.22 and sim. in Cond. SMB 36, p. 46.21).³⁵ These occurrences must be distinguished from the clearly Romance expressions like *om(n)i annu* ‘every year’ (StSS L. I-XIX, p. 10.47 and StSS L. I-CVIII Rubr., p. 46.14, StSS L. I-CXLIX, p. 59.12; sim. in Cond. SPS 426, p. 342.20), which might be Italianisms (from it. *ogni anno*, see below), as was suggested by Wagner (DES 2: 188), with a Latinizing spelling. Finally, there is one isolated case with the parasitic *d-* that we have seen above: *p(er) don(n)j bolta* (CdLA CI, p. 140.23).

Ogni is a clear Italianism, which exclusively appears in the *Statuti Sassaresi* and the *Statuti di Castelsardo*. Some examples are *per ogni cavallu* (‘for each horse’, StSS L. II-XLVIII, p. 82.38), *de ogni atheru po(r)chu* (‘of every other pig’, StSS L. II-L, p. 83.4), and the time expressions³⁶ *ogni annu* (‘every year’, several times, e.g. in StSS L. II-XLVI, p. 82.15), *ogni die* (‘every day’, StCastel. LVIII, p. 31.23). Like in Italian, this item is only used with singular nouns and does not exist as a pronoun. Similarly to *ogna*, the *Statuti Sassaresi* show three occurrences with initial *d-* (*dogni*), a blend of It. *ogni* and the probably autochthonous *donnia* (see above).

³⁴But see StSS 80,40 with the Italianism *ogni* and another syntax: *ogni inpedimentu cessante* ‘every obstacle having ceased to exist’.

³⁵The sense of Cond. SMB 36, 46.18 is not totally clear to me:

(i) E de omni apat fine a su fine in seculum.
 and of all have.SUBJV-3SG end to the end in saeculum

Viridis (2003: 113) translates “E tutto ciò in perpetuo.” (‘And all this forever.’)

³⁶See the similar use of *ogna* in footnote 30.

Finally, the corpus shows a total of 45 occurrences of *omnes* and its variant *onnes* (< Lat. OMNES). As far as we can judge from the texts that can be dated more or less precisely, this item ceased to be used at the beginning of the 13th century. However, a productive use of *omnes* is only found three times in the *Privilegio logudorese* (1080–1085):

- (21) a. (Priv. Log., p. 253.2)
Ego iudice Mariano de Lacon, fazo ista(m) carta ad onore de **om(ne)s homines** de Pisas
‘I, the judge Mariano de Lacon, make this document for the honor of all men of Pisa’
- b. (Priv. Log., p. 254.9–10)
ego feci-nde-lis carta pro honore de xu piscopu(m) Gelardu e de Ocu Biscomte e de **om(ne)s consolos** de Pisas e ffeci-la pro honore de **om(ne)s ammicos meos** de Pisas: [...]
‘I make this document for the honor of the bishop Gelardu, and Ocu Biscomte, and all consules of Pisa, and I make it for the honor of all my friends from Pisa: [...]’

We cannot tell whether this is evidence for an earlier use of *omnes* in Sardinian that had become almost obsolete at the beginning of the written documentation or whether the writer(s) of the *Privilegio logudorese* used *omnes* as a Latinism. In any case, in other 11th century texts, even of a slightly earlier date, *omnes* only appears followed by the adjective or noun *sanctu* (Sard.)/*sanctus* (Lat.) ‘holy, Saint’ and preceded by *i(n) grat(tia) de* ‘in thanks to’, as in *i(n) grat(tia) de [...]* *o(mn)es s(anc)tos P(ro)ph(et)as* ‘in thanks to [...] all holy prophets’, CVolg. AAC 1, p. 43, 1.16), *i(n) grat(tia) de [...]* *o(mne)s s(anc)ti Martires* ‘in thanks to [...] all holy martyrs’, with the Latin nominative plural *sancti*, CVolg. AAC 1, p. 43, 1.21), and *i(n) grat(tia) de [...]* *o(mne)s s(anc)tos et s(anc)tas Dei* ‘in thanks to [...] all saints of God’, CVolg. AAC 1, p. 43, 1.22). In all the later documentation, too, *omnes* (and 4 occurrences of *onnes*, all stemming from the Sardinian documents of the monastery of Montecassino)³⁷ is only found in such fixed formulaic expressions, mostly in strongly Latinizing (parts of) texts. The relevant expressions are almost exclusively two formulae. The first is the formula that we have already seen, *omnes/onnes sanctos et sanctas Dei*, the second *omnes/onnes frates meos e fideles meos testes* ‘with all my brothers and my stalwarts as witnesses’.

³⁷For the tight relationship of the monastery (located on the Italian mainland) to Sardinia and its activities on the island as well as the documents at issue (mostly donation letters, see Saba 1927).

To summarize, we can say that Latin OMNIS yielded the universal quantifier (*d*)*omnia* (from the Latin neuter plural OMNIA) as an inherited word, which is found in two functions: first, as a pronoun meaning ‘all, everything’, which must, however, be restricted by a relative clause; second, like some Italianisms also derived from OMNIS, for quantifying over an individual expressed by a singular NP (‘every/each X’). For the plural, we find a small number of early occurrences of *omnes*. However, ‘all.PL’ was mostly expressed by *tot(t)u*, as we shall see in the following subsection.

4.3 *Tot(t)u* ‘all’

The item *tot(t)u* stems from late Latin TŌTTUS, a variant of TŌTUS (REW 8815, DES 2: 500). Its main functions are universal plural quantification like in (22) and universal quantification of singular mass, collective and abstract nouns as shown in (23).

- (22) a. (Cond. SMB 209, p. 134.21)
denanti dess’altari suo, ue erant **totu sos monagos**
‘in front of his altar, where all the monks were’
- b. (CVolg. AAC 1, p. 43, 2.44)
a ponner curadores et maiores suos i(n) **totas billas** dessus paniliu<s>.
‘to put officials and principals of his in all villages of semifree serfs.’
- (23) a. (CVolg. AAC 1, p. 43, 2.10)
Et **totu custu serbiciu** fage(n)ta fina ad icomo ad su Re(n)nu.
‘And they have been doing all this service to the Kingdom until now.’
- b. (Carta Mars. 2, p. 72, 2.12)
(et) de **totu bi[l]la** de Maara
‘and of all the village of Maara’
- c. (StSS L. I-XLIII, p. 23.42)
q(ui) **totta s’abba** de cussas co(n)ças se vochet foras dessa terra de Sass(ar)i
‘that all the (waste)water of these tanneries should be poured away outside the territory of Sassari’

Let us consider some aspects of the syntax of Old Sardinian *tot(t)u* (deferring the lack of agreement in examples such as (22a) and (23b) until §5). Unlike modern Sardinian, the determiner following *tot(t)u* (either a definite article as in (22a) and (23c) or a demonstrative as in (23a)) was not obligatory, see (22b) and (23b). Occasionally, *tot(t)u* occurs to the right of the NP or DP, e.g. *fios suos tottu* ‘all

her children' (Cond. SPS 205, p. 210.17) besides *cu(n) tottu fijos suos* 'with all her children' (ibidem, p. 210.33). As both examples are found in the same context (a long list of names of freed serfs), there does not seem to be any semantic or pragmatic difference between the two. Strikingly, postposed *tot(t)u* mostly appears when a relative clause or some restricting phrase follows:

- (24) a. (CVolg. AAC 20, p. 315.16)
 Et dau illoy **su saltu miu de genna de Codrigla totu** in qua si segat.
 'And I give him my all my (woodland) pasture of Genna de Codrigla up to where it is delimited.'
- b. (Cond. SMB 32, p. 40.16)
 et **fundamentu suu totu** c'aviat in Calcaria de Comita de Muru
 'and all his land that he had in Calcaria de Comita de Muru'
- c. (Cond. SMB 202, p. 127.13)
 et **ipsa binia sua tota** de Tommanu
 'and all his vineyard of Tommanu'
- d. (StSS L. I-X, 7.38)
Sos bandos tottu in custu breve c(on)tentos, missos (et) ma(n)datos
p(er) issu
 'All bans contained, issued and authorized by him'

I provisionally interpret these examples as structures in which the property that determines the set expressed by 'all' is spelled out right adjacent to the quantifier. Since this is a phrasal constituent or, in generative terms, a maximal projection, it cannot be inserted in the standard head position (Q^0) of a quantifier phrase and it must therefore be generated in a right peripheral position.³⁸

Instead of being part of a NP or DP, *tot(t)u* could be used predicatively, with the meaning 'entirely'. As (25b) shows, this sense could additionally be made explicit by the item *intre(g)u* 'entire':

- (25) a. (CVolg. AAC 9, p. 63, 2.11)
 una domu **totu** fabrigada (et) cob(er)ta
 'a house entirely built and covered'

³⁸It is striking that there is a possessive adjective in (24a–c). However, the existence of the possessive did not obligatorily trigger postnominal *tot(t)u*, as the following example shows: **totu sa t(er)ra n(ost)ra de Caralis** 'all our land of Cagliari' (CVolg. AAC 1, p. 44, 1.30). It seems that the structure seen in (24a–d) has a kind of partitive meaning, e.g. for (24b): 'the complete part of his land that he had in Calcaria de Comita de Muru'. Thanks to Olga Kellert for pointing this out to me.

- b. (Cond. SMB 28, p. 35.25)
 et Iohanne de Urri ramasit a *sanctu* Georgii **totu** intreu
 ‘and Iohanne of Urri remained entirely in the possession of St.
 George’

Examples like those in (26) are similar, but unlike those in (25), they can be seen as cases of quantifier floating.

- (26) a. (Cond. SMB 26, p. 34.13)
 Et sa parte de sa mugiere, si obierit sine filiis, remaneat **tota** assa
 domo de *sancta* Maria *pro s’anima* sua.
 ‘And the wife’s part, if she dies without children, shall all be left to
 Saint Mary for (the well-being of) her soul.’
 b. (CdLA VI, p. 60.25)
 et issos b(e)nis suos **tottu** siant (con)flischados assa co(r)ti n(ost)ra
 ‘and their possessions shall all be confiscated by our court’
 c. (Carta Mars. 2, p. 72, 1.20)
 ca fuit **totu** de S(an)c[tu Satur]ru su saltu
 ‘because the (woodland) pasture belonged all to Saint Saturru’

Of the 545 occurrences of *tot(t)u* in the corpus,³⁹ only around 20 cases are of the types in (25) and (26).⁴⁰ In another 35 cases, *tot(t)u* is clearly used as a pronoun. In contrast to adnominal *tot(t)u*, whose documentation begins in the 11th century, the pronominal use does not seem to be attested earlier than the beginning of the 13th century.⁴¹ Usually, pronominal *tot(t)u* is uninflected and is used either as a singular (‘all, everything’ as in (27a–c) (24 cases) or as a plural (‘all [of them]’) as in (27d,e).⁴² The singular *tot(t)u* clearly competes with *omnia* and *donia* (see §4.2), but unlike the latter normally occurs without a restricting relative clause (except for some rare cases such as (27c).⁴³

³⁹Among these, I counted one occurrence of *tutu* and one of *tuta*, where the *-u-* is probably an Italian influence. I did not count four occurrences of the plural form *tuti*, which is clearly Italian.

⁴⁰Also including a small number of cases in which *tot(t)u* is used predicatively with an empty subject (pro or PRO) as an antecedent.

⁴¹With the caveat that there are three occurrences in the *Condaghe di San Pietro di Silki* (end of 11th–13th c.), which cannot be dated exactly.

⁴²As for the plural reading ‘all (of them)’, the inflected form *totos* is only documented twice as a pronoun in the whole corpus (Cond. SMB 131, 86.6; Cond. SMB 1, 8.10).

⁴³In addition, we find *tottu (is)su chi* ... ‘all this that ...’, but here *tot(t)u* is adnominal and not a pronoun.

- (27) a. (Cond. SMB 23, p. 30.12)
Totu lu damus a *sancta* Maria de Bonarcadu prossas *animas nostras*.
 ‘We give all to Saint Mary of Bonarcado for (the well-being of) our souls.’
- b. (Cond. SMB 66, p. 54.13)
 et est **totu** puspape .XXX. sollos
 ‘and it is all together (worth) thirty soldos’
- c. (CVolg. AAC 11, p. 294.28)
 Istimo(n)ius [...] de **totu** ca(n)tu narat ista carta, do(n)nu Riccu su archipiscobu miu de Pluminus, et [...] ‘Witnesses [...] of all that is said in this document [are] Donnu Riccu, my Archbishop of Pluminus, and [...]’
- d. (StSS L. I-I, p. 4.21)
 Iustithia aces facher ad **tottu**, man(n)os et piçinnos
 ‘You have to do justice to all (of them), adults and children’
- e. (CdLA XVI, p. 72.11)
 si **totu** o sa maiore parte non esserent in concordia no siant credidos
 ‘if all (of them) or the majority do not accord, they shall not be believed’

In the *Statuti Sassaresi*, the singular reading also appears in formulaic expressions containing *in tottu*, such as *in tottu et per tottu* ‘in all and for all’ (i.e. entirely, in all respects) and *in tottu over in parte* ‘totally or partially’, which are calques from the equivalent Italian expressions *in tutto e per tutto* and *in tutto ovver’ in parte*.

5 Agreement patterns of Old Sardinian *tottu*+DP/NP

As mentioned in §4.3, adnominal *tottu* often appears as invariable in Old Sardinian, i.e. without agreement with the DP or NP. Interestingly, structures with and without agreement of *tot(t)u* can be found in the corpus (contra Blasco Ferrer’s 1984: 93 observations).⁴⁴ In fact, both structures, with and without agreement of *tot(t)u*, are documented in the texts:

⁴⁴ [Il lat. volg. TOTUS] [...] si è cristallizzato sin dalle prime documentazioni nella forma invariabile /tóttu/”.

- (28) a. fem. sg. [+agreement] (Cond. SPS 44, p. 118.7)
e llevarun **totta** sa casa issoro
'and they took away all their possessions'
- b. fem. sg. [-agreement] (Cond. SMB 104, p. 74.19)
Parsit iustitia a **totu** corona de logu
'It seemed just to the whole court'
- (29) a. fem. pl. [+agreement] (CdLA CXXV, p. 166.3)
sas dominiguas de totu s'an(n)o et **totas** sas festas de santa Maria
'the Sundays of the whole year and all the feasts of Saint Mary'
- b. fem. pl. [-agreement] (StSS L. I-XXXVII, p. 21.21)
deppiat satisfacher sa mesitate d(e) **tottu** sas ispesas
'[he] had to cover half of all the expenses'
- (30) a. masc. pl. [+agreement] (Cond. SMB 33, p. 41.2)
ad honore de Deus et de sancta Maria et de **totos** sos santos
'to the honor of God and of Saint Mary and of all the saints'
- b. masc. pl. [-agreement] (Cond. SNT 1, p. 64.14)
Testes: Simio d'Elices e **totu** bicanos suos.
'Witnesses: Simio d'Elices and all his neighbours'

Let us look at the distribution of agreeing and non-agreeing adnominal *tot(t)u*, shown in Table 5.

If we first look at the last line, we see that, very strikingly, *tot(t)u* agreed in the overwhelming majority of cases (89%) in the feminine singular, whereas it rather rarely agreed in the feminine plural (ca. 12%) and not very frequently either in the masculine plural (ca. 19%). Due to the fact that the *condaghes* (Cond. SPS, Cond. SMB, Cond. SNT) contain texts that extend over great time-spans, it is rather difficult to make any statement concerning the diachronic development. We can however say that, at least by tendency, agreement in the feminine singular seems to have been constantly predominant (ca. 90%–100%) until the end of the 14th century, when it suddenly drops to 50% in the latest text. The masculine plural form seems to have had at least some significant vitality between the end of the 11th and the second half of the 13th c. and was practically inexistent in the 14th c. Together with the drop of the frequency of the feminine singular form, we could interpret this as the beginning of a tendency that would ultimately lead to the modern situation without agreement. There are, however, two issues that have to be addressed concerning the plural forms.

First, as we have seen, the feminine plural form is scarcely found throughout most of the documentation, but strikingly, the latest text of the end of the 14th

Table 5: Agreement of adnominal *tot(t)u*

Texts/ Period	Singular			Plural				Total
	m. <i>tottu</i>	f.		m.		f.		
		+agr	-agr	+agr	-agr	+agr	-agr	
		<i>totta</i>	<i>tottu</i>	<i>tottos</i>	<i>tottu</i>	<i>tottas</i>	<i>tottu</i>	
div. texts (1050–1150)	4	8 (88.9%)	1 (11.1%)	1 (20%)	4 (80%)	1 (100%)	0 (0%)	19
div. texts (1150–1225)	16	1 (50%)	1 (50%)	1 (9.1%)	10 (90.9%)	0 (0%)	7 (100%)	36
Cond. SPS (end 11 th –mid 13 th c.)	35	34 (100%)	0 (0%)	5 (50%)	10 (50%)	0 (0%)	2 (100%)	86
Cond. SMB (12 th – 13 th c.)	26	39 (84.8%)	7 (15.2%)	8 (40%)	12 (60%)	0 (0%)	0 (0%)	92
Cond. SNT (1 st quarter 12 th –2nd half 13 th c.)	9	22 (100%)	0 (0%)	13 (62%)	9 (38%)	2 (100%)	0 (0%)	55
StSS (1316)	24	10 (90.9%)	1 (9.1%)	0 (0%)	40 (40%)	0 (0%)	33 (100%)	108
StCastel. (1334–1336?)	2	2 (100%)	0 (0%)	0 (100%)	16 (100%)	0 (0%)	14 (100%)	34
CdLA (end 14 th c.)	7	5 (50%)	5 (50%)	1 (4%)	24 (96%)	5 (71.4%)	2 (28.6%)	49
Total	123	121 (89%)	15 (11%)	29 (18.8%)	125 (81.2%)	8 (12.1%)	58 (87.9%)	479

century (the *Carta de Logu*) shows 5 occurrences of this form (*totas*) vs. 2 of the non-agreeing form (*tot[t]u*) in feminine plural contexts.⁴⁵ We have already seen in previous sections that the *Carta de Logu* is among those texts that are particularly prone to Italianisms. However, it would not be plausible to consider the form *totas* or at least the tendency to have agreement in the feminine plural as an Italianism, as the other strongly Italianizing texts (*Statuti Sassaresi*, *Statuti di Castelsardo*) do not show this phenomenon, i.e., in these texts, *tot(t)u* never agrees in the plural. The only reasonable conclusion seems to me to consider this as a Catalanism, also taking into account that the corresponding Old Catalan form was actually *totas*.⁴⁶

Second, when we look at the occurrences of the masculine plural form (*tot([t]os)*, we observe that, interestingly, 22 of the 29 occurrences are all followed by a numeral. More particularly, most of these occurrences show a structure of the type ‘*all (the) numeral N*’, as shown in (31):

- (31) a. (Cond. SPS 33, p. 108.36)
torraitimilos iudike **tottos** .VI. **sos fijos** de Barbara Rasa
‘the judge gave back to me all six sons of Barbara Rasa’
- b. (Cond. SNT 1, p. 107.2)
et a Petru de Nurki et a **totos** .III. **sos connatos** comporailis su pede
de Iorgi de Contra
‘and from Petru de Nurki and all his three brothers-in-law, I bought a
quarter of Iorgi de Contra’
- c. (Cond. SMB 133, p. 89.9)
Mandei pro·llos et benneruntimi **totos tres frates fijos** de Gostantine
Stapu: Orçoco et Comida et Iohanne.
‘I summoned them, and there came all three brothers, sons of
Gostantine Stapu: Orçoco, Comida, and Iohanne.’

⁴⁵This unexpected rise is only partially explained because three of the five occurrences of the agreeing forms stem from the same passage containing three times the same pattern:

- (i) (CdLA CXXV, p. 166.3–4)
In p(ri)mis sas dominiguas de totu s’an(n)o et **totas sas festas** de santa Maria; item
totas <sas> festas de sos apostollos e **tot(a)s sas festas** de sos evangellistes; [...]’
‘First, the Sundays of the whole year and all the feasts of Saint Mary, then all the feasts
of the Apostles, and all the feasts of the Evangelists; [...]’

⁴⁶The O. Cat. paradigm of this quantifier was *tot* (m. sg.), *tota* (f. sg.), *totz* (m. pl.), *totas* (f. pl.). For Catalan influences in the *Carta de Logu*, see Loi Corvetto (1992: 180–181), where the author discusses some ideas by Sanna (1975: 136). Even if this author thinks that some rather clear Catalanisms such as *desviadu*, *mescladura*, *biage* (var. *biagio*, *biatgio*) and the spellings *que-*, *gue-* for It. *che-*, *ghe-* might be explained otherwise (e.g. as influences of Genovese), he admits the possibility of an Aragonese scribe having been involved in the writing of the manuscript.

Conversely, lack of agreement of *tot(t)u* is not found at all whenever a numeral follows. This even turns out to be true for the feminine, where only one occurrence with a numeral is found in the corpus, showing agreement:

(32) (Cond. SNT 1, 94.5)

Conporailis ad Ytçoccor Mavronti et assos *frates*, die de Pale Pirinione, et die in Istefane Pira, et .iii. dies in *totas .iii. sas filias*: [...]

‘I bought from Ytçoccor Mavronti and from his brothers one day of Pale Pirinione, and one day of Istefane Pira, and three days of all three daughters: [...]’

We can thus summarize the results as follows: agreement of *tot(t)u* was optional in Old Sardinian, with a strong preference towards agreement in the singular. In the plural, agreement is only marginally attested but was obligatory when *tottu* was followed by a numeral.

Some more comments can be made with respect to the structures in (31) and (32), which are interesting from a typological perspective. Within the modern Romance languages, there are basically two patterns, which are shown for Italian in (34): the structure illustrated in (34a), with the conjunction *e* ‘and’ before the article (situated between the numeral and the noun), and the option in (34b), without the conjunction and the article preceding the sequence numeral + noun. Whereas French does not allow at all the combination of *tous* ‘all’ with a numeral (cf. Doetjes 1997: 210, see ex. (33)), Spanish only allows the option corresponding to (35b), whereas (35a) is ungrammatical:

(33) French

- a. *Tous (et) trois les étudiants ont lu le livre.
- b. *Tous les trois étudiants ont lu le livre.
‘All three students read the book.’

(34) Italian (cf. Balsadella 2017: 7)

- a. Tutti e tre gli studenti hanno letto il libro.
- b. Tutti i tre studenti hanno letto il libro.
‘All three students read the book.’

(35) Spanish (cf. Cirillo 2009: 173)

- a. *Todos y tres (los) estudiantes leyeron el libro.
- b. Todos los tres estudiantes leyeron el libro.
‘All three students read the book.’

In contrast, Old Sardinian had another structure, as witnessed in (31a,b) to (32), which was similar to (34a) but lacking the coordinating conjunction and with the definite article being optional (see 31c). The same word order can be shown to have existed in other medieval Romance languages:

- (36) Middle Italian (OVI, *Bibbia volgare* Ez 41, p. g575)
e due porte erano da **tutti due li lati** degli usci
'and two doors were at both two sides of the exits'
- (37) Old Spanish (CORDE, Anónimo, c1414)
todos quatro los caualleros mobieron
'all three knights moved on'
- (38) Old French (*Guillaume d'Orange*, v. 792, ed. Jonckbloet 1854: 134)
Quant il connut **toz trois les compaignons**
'When he recognized all three companions'

What distinguishes Old Sardinian from these languages is that agreement on the universal quantifier was optional and becomes obligatory exactly in this structure involving a numeral. On a more typological level, outside the Romance languages, the Old Sardinian structure is identical to the option (39b) of Modern Dutch. Interestingly, we find a similar agreement pattern (with agreeing *all*, in contrast to 39a):

- (39) Modern Dutch (Cirillo 2009: 160)
- a. Al de drie studenten hebben het boek gelezen.
all the three student-PL have the book read
- b. Alle drie de studenten hebben het boek gelezen.
all-PL three the student-PL have the book read

Cirillo (2009: 160) analyzed these structures as follows:

- (40) a. [QP all [DP the [CardP three [NP students]]]]
b. [QP all three [DP the [CardP $\emptyset$ [NP students]]]]

In (40a), which corresponds to the word order in (39a) as well as to that of the Italian structure in (34b) and the Spanish structure in (35b), the quantifier is generated in its regular position within a quantifier phrase (QP) preceding the DP, and the cardinal number is generated beneath the determiner in a cardinal phrase (CardP). (40b) represents the word order of the Dutch example (39b), which is the same as in the Old Sardinian examples in (31) and (32) and of the other Old Romance languages shown in (36) to (38). Here, in Cirillo's analysis, the numeral

is generated together with the quantifier (creating a complex quantifier head, the “universal numeric quantifier ($\forall\text{NumQ}$)”,⁴⁷ Cirillo does not account for the agreement behavior, yet. An explanation for the presence of agreement in (39b) is found in Corver (2010): According to him, the $\forall\text{NumQ}$ expression *alle drie* ‘all three’ is also generated as a complex head (Num°).⁴⁸ The derivation is performed in two steps: first, the NP is raised to the specifier of NumP , and this is where agreement is realized, via specifier head agreement. Then, the lower part of the NumP moves to the specifier of DP.⁴⁹

I have included this brief discussion because it shows that the Old Sardinian data can contribute to some interesting issues of a general linguistic interest, but I will not go any further into the formal analysis of this construction. Let us, instead, have a brief look at Modern Sardinian. Recall from §2.2 that despite the general tendency of Modern Sardinian *tot(t)u* to show lack of agreement, a modern plural form *tot(t)us* can be observed in some syntactic contexts. First of all, as I already said in §2.2, this form is striking, because in the modern Logudorese and Nuorese varieties, this form with the ending *-us* cannot be derived from the old plural form in *-os*, as these varieties preserve *-o* in word-final syllables. In addition, as we have seen, the old plural forms were almost not used at the end of the medieval period and, in any case, when they occur, there was a clear distinction between a masculine form ending in *-os* and a feminine form ending in *-as*, whereas the modern plural form in *-us* is invariable for gender. It is therefore very probable that the modern form *tot(t)us* is an innovation. Why this innovation arose in contexts such as that in example (3) of §2.2 must be left for future research. However, Jones (1993: 38) mentions another syntactic context in which some speakers use the plural form, namely in connection with numerals:

- (41) *tottus tres ómines*
 all-PL three men
 ‘all three men’

⁴⁷Note that he does not derive (40b) from (40a) because head-movement from the lower Card-Position would have to cross D° and thus violate the HMC. The structure in (40b) is, however, somewhat awkward because of the empty Card head.

⁴⁸The NumP has the same position of Cirillo’s CardP: [NumP [Num° all five] [NP women]].

⁴⁹This step is problematic because it involves movement of an X' -constituent, which should not be allowed in modern generative frameworks. In Mensching (2023) I discuss Corver’s (2010) assumptions in more detail and suggest a similar analysis (applied to the Old Sardinian structures) that does not have this problem. I propose that the NP moves out of the CardP or NumP to a functional category, and the remaining part of NumP/CardP (containing the $\forall\text{NumQ}$ item) undergoes remnant movement to QP. Obligatory agreement is explained withing the minimalist framework following Chomsky (2000): The Q head has an unvalued phi-probe that probes the NumP remnant, and thus needs valued phi features on $\forall\text{NumQ}$.

Note that, if this is a $\forall$ NumQ construction, agreement is expected (once the language has a plural form), if an analysis along the lines of Corver (2010) is on the right track. Unfortunately, since Jones's examples lack the article, we cannot exactly determine its structure.⁵⁰

6 Summary and outlook

In the present chapter, I have presented some preliminary results of a corpus analysis on indefinites in Old Sardinian, a language that had been understudied in this regard. My interest in such a study was mostly motivated by (i.) the fact that a part of the inventory of indefinites of Modern Sardinian is known to contain indefinites that are loanwords mostly taken from older stages of Italian, Catalan, and Spanish, and (ii.), that some Modern Sardinian quantifiers show lack of agreement. In both cases, for Old Sardinian, i.e. the language documented in medieval legal and administrative documents, these issues had been considered before in the literature, but only on a superficial level, which had led to hypotheses that had never been matched against quantitative data. With respect to (i.), the quantitative methods applied here (on the basis of the corpus ATLiSOr) led to results that can be summarized as follows:

1. Wagner (1938–1939) and in his DES tended to consider the negative indefinites *nullu*, *perunu* 'no (X)' *neunu*, and *nexunu* 'nobody/no (X)' (and their variants) as Italianisms. I have shown that the distribution of these items over time strongly suggests that only the latter two are loans from Old Italian. Whereas *nullu* should be considered either a Latinism or an item inherited from Latin (maybe rather the former given its scarce documentation), *perunu* must definitely be interpreted as an autochthonous element derived from Latin.

⁵⁰The Modern Sardinian structures need more research. An informal inquiry that I made with three speakers of Logudorese and Nuorese varieties suggests that, when the article is present, the following options are possible:

- (i) a. totu(s) e tres sos òmines.
all(-PL) and three the men
- b. totu sos tres òmines.
all the three men

These options correspond to the Italian patterns in (34a,b) and are therefore certainly to be considered as Italianisms, particularly because they are not found in the medieval documentation. Interestingly (and coherent with what we have said about $\forall$ NumQ-constructions), only (i.a) optionally appears to admit agreement for some speakers.

2. As for *cada* ('every/each'), Wagner (DES 1: 256) used the Old Sardinian documentation for arguing that this item is probably not a Hispanicism. However, when looking at the corpus, it becomes evident that *cada* does not appear at all in the corpus in its modern sense. Apart from one occurrence with a distributive sense of 'X N each' (with X a numeral), this item is only found in the lexicalized compound *cadadie* 'daily' in expressions meaning 'full-time serf/maid'. I therefore concluded that *cada* in the sense of 'every/each' must be a later loan from Catalan and Spanish.
3. A small number of early occurrences of *omnes/onnes* 'all.PL' might indicate that this item was inherited from Latin but was almost obsolete at the beginning of the Old Sardinian documentation. In any case, later occurrences of these items are clear Latinisms that are only found in Latinizing texts and formulae.
4. My analysis strongly suggests that the state of the art concerning *omnia* and *onna* 'every/each' must be revised as follows: they are neither Latinisms nor Italianisms. I rather consider *onna* as a form inherited from OMNIA in its Vulgar Latin singular use and *omnia* as a Latinizing spelling variant. Wagner's idea that *omnia* and *onna* are spelling variants of the Italianism *ogna* is contradicted by the fact that *ogna* appears rather late in some Old Sardinian texts that are known for their Italianizing tendencies and that neither *omnia* nor *onna* were usual in Old Italian.

Apart from these findings, the corpus study also provided an occasion to look at some aspects of the syntax of these elements. In this respect, an innovative finding is that Old Sardinian seems to have been a strict negative concord language, unlike Modern Sardinian, where preverbal negative indefinites lack negative concord. Contrarily to the items *nullu* and *perunu*, which I have argued to be autochthonous elements, the borrowed negative indefinites often show the lack of negative concord when they appear preverbally, which seems to indicate that this property has been adopted from Old Italian together with the items itself. More generally, the occurrence of these items in preverbal positions also indicates foreign influence, given that Old Sardinian was mostly a V1 language, a fact reflected quite well in the syntax of *perunu* and *nullu*.

Finally, I have been looking at the agreement behavior of *tot(t)u* 'all, whole', inherited from late Latin TŌTTUS. In modern Sardinian, this item standardly lacks agreement. For Old Sardinian, an in-depth study of this item was missing until now and has been provided in this article for the first time. The results show that agreement of *tot(t)u* was optional in Old Sardinian, with a strong preference for agreement in gender in the singular (*tot(t)u* vs. *tot(t)a*), whereas, in the

plural ((*tot(t)os/tot(t)as*) agreement in both genders was strongly dispreferred (in favor of the default form *tot(t)u*). The agreement property of this item must have vanished altogether after the Middle Ages, and the gender-neutral plural form *tot(t)us*, which is occasionally observed in some modern varieties, must be considered as an innovation. I have also been able to detect an exception to the optional agreement of Old Sardinian *tot(t)u*, namely a structure involving numerals of the type ‘*all* Numeral (Det) N’ (in this word order), where I have identified the sequence ‘*all* Numeral’ as an instance of universal numeric quantifiers according to Cirillo (2009). Here, number agreement in the plural was obligatory, in conformity with observations that have been made for other languages which show this kind of structure.

All in all, I hope to have demonstrated with this study that a thorough corpus-based analysis on Old Sardinian can bring forth important insights in the field of indefinites, both for Romance and for general linguistics.

Acknowledgments

This chapter was written within the framework of a joint project with Cecilia Poletto (“Quantification in Old Italian”) financed by the Deutsche Forschungsgemeinschaft (DFG).

Sources and dictionaries

Carta arb.	Carta arborenses del 1184. In Eduardo Blasco Ferrer (ed.), <i>Crestomazia sarda dei primi secoli</i> (2003), vol. 1, 125–128. Nuoro: Ilisso.
Carta arb. Gen. 2	Seconda carta arborenses di Genova. In Eduardo Blasco Ferrer (ed.), <i>Crestomazia sarda dei primi secoli</i> (2003), vol. 1, 104–108. Nuoro: Ilisso. Manuscript dated 1112–1120.
Carta Ben.	Carta di Benedetta de Lacon. In Eduardo Blasco Ferrer (ed.), <i>Crestomazia sarda dei primi secoli</i> (2003), vol. 1, 93–96. Nuoro: Ilisso. Manuscript dated 1225.
Carta don.	Carta di donazione di Guglielmo-Salusio. In Eduardo Blasco Ferrer (ed.), <i>Crestomazia sarda dei primi secoli</i> (2003), vol. 1, 85–88. Nuoro: Ilisso. Manuscript dated 1211.
Carta gall.	Carta gallurese. In Eduardo Blasco Ferrer (ed.), <i>Crestomazia sarda dei primi secoli</i> (2003), vol. 1, 177–183. Nuoro: Ilisso. Manuscript dated 1173.

- CdLA Giovanni Lupinu (ed.). 2010. *Carta de Logu dell'Arborea: Nuova edizione critica secondo il manoscritto di Cagliari (BUC 211) con traduzione italiana*. Oristano: Istituto Storico Arborense.
- Cond. Bar. II Giovanni Lupinu & Sara Ravani. 2015. Per una nuova edizione critica del "Condaghe di Barisone II". *L'Italia Dialettale* 73. 49–74.
- Cond. SMB Maurizio Virdis (ed.). 2002. *Il Condaghe di Santa Maria di Bonarcado*. Cagliari: CUEC.
- Cond. SNT Paolo Merci (ed.). 1992. *Il condaghe di San Nicola di Trullas*. Sassari: Carlo Delfino.
- Cond. SPS Alessandro Soddu & Giovanni Strinna (eds.). 2013. *Il condaghe di San Pietro di Silki*. Nuoro: Ilisso.
- CVolg. AAC Arrigo Solmi. 1905. Le carte volgari dell'Archivio Arcivescovile di Cagliari: Testi campidanesi dei secoli XI–XIII. *Archivio Storico Italiano* 35. 273–330.
- CVolg. AAC 1 Carta volgare dell'Archivio Arcivescovile di Cagliari n. 1. In Eduardo Blasco Ferrer (ed.), *Crestomazia sarda dei primi secoli* (2003), vol. 1, 43–50. Nuoro: Ilisso. Manuscript dated 1066–1074.
- DES Max Leopold Wagner. 1960–1964. *Dizionario etimologico sardo*. Heidelberg: Winter.
- FEW Walther von Wartburg et al. 1922–2002. *Französisches Etymologisches Wörterbuch: Eine Darstellung des galloromanischen Sprachschatzes*. Bonn, Heidelberg, Leipzig-Berlin, Basel: Klopp/Winter/Teubner/Zbinden. 25 vols.
- Montecass. 5 Giovanni Lupinu (ed.). *Carta di donazione di Gonario de Laccon ed Elene de Thori a Montecassino*. Edizione a uso interno.
- Montecass. 9 Giovanni Lupinu (ed.). *Carta di donazione di Costantino de Carbia e Giorgia de zZori a Montecassino*. Edizione a uso interno.
- Montecass. 10 Giovanni Lupinu (ed.). *Carta di donazione di Comita de Azzen e Muscunione de zZori a Montecassino*. Edizione a uso interno.
- Montecass. 12 Giovanni Lupinu (ed.). *Carta di donazione di Furatu de Gitil e Susanna de zZori a Montecassino*. Edizione a uso interno.
- Montecass. 16 Carta di donazione di Furatu de Gitil a Montecassino. In Eduardo Blasco Ferrer (ed.), *Crestomazia sarda dei primi secoli* (2003), vol. 1, 160–164. Nuoro: Ilisso.

- Montecass. 20 Giovanni Lupinu (ed.). *Carta dell'abate Benedetto*. Edizione a uso interno.
- Montecass. 22 Carta di donazione di Costantino de Athen a Montecassino. In Eduardo Blasco Ferrer (ed.), *Crestomazia sarda dei primi secoli* (2003), vol. 1, 170–173. Nuoro: Ilisso.
- Montecass. 32-orig Giovanni Lupinu (ed.). *Carta di conferma e di concessione di Gonario de Laccon*. Edizione a uso interno.
- Montecass. 35 Carta di revoca tributaria a favore di Montecassino. In Eduardo Blasco Ferrer (ed.), *Crestomazia sarda dei primi secoli* (2003), vol. 1, 174–176. Nuoro: Ilisso.
- Montecass. 39 Giovanni Lupinu (ed.). *Carta di Barisone d'Arborea in favore di Montecassino*. Edizione a uso interno.
- Priv. Log. Privilegio Logudorese. 1986. In Santorre Debenedetti (ed.), *Studi filologici. Con una nota di Cesare Segre*, 248–261. Milan: Franco Angeli. Manuscript dated 1080–1085.
- REW Wilhelm Meyer-Lübke. 1992. *Romanisches etymologisches Wörterbuch*. 6th edn. Heidelberg: Winter.
- StCastel. Statuto di Castelsardo. 1899. In Enrico Besta (ed.), *Intorno ad alcuni frammenti di un antico statuto di Castelsardo. Archivio giuridico "Filippo Serafini"* nuova serie 3.2, 231–332.
- StCastel. 2 Statuti di Castelsardo (capp. 190–199). In Eduardo Blasco Ferrer (ed.), *Crestomazia sarda dei primi secoli*, vol. 1, 189–192. Nuoro: Ilisso. Manuscript dated 1334–1336.
- StSS Pier Enea Guarnerio. 1892–1894. Gli statuti della Repubblica Sassarese, testo logudorese del secolo XIV, nuovamente edito d'in sul codice. *Archivio glottologico italiano* 13. 1–124.
- Tratt. Pace Trattato di pace del 1206. In Eduardo Blasco Ferrer (ed.), *Crestomazia sarda dei primi secoli*, vol. 1, 77–84. Nuoro: Ilisso.

References

- Balsadella, Davide. 2017. *I tipi sintattici "tutti (e) tre" e "tutti i tre" in italiano. Uno studio diacronico*. Università Ca' Foscari Venezia. (MA thesis). <http://dspace.unive.it/bitstream/handle/10579/10552/837261-1203969.pdf> (14 June, 2021).
- Beghelli, Filippo & Tim Stowell. 1997. Distributivity and negation: The syntax of each and every. In Anna Szabolcsi (ed.), *Ways of scope taking*, 71–108. Dordrecht: Kluwer.

- Bertocchi, Alessandra, Mirca Maraldi & Anna Orlandini. 2010. Quantification. In Philip Baldi & Pierluigi Cuzzolin (eds.), *New perspectives on historical Latin syntax*, vol. 3, 19–174. Berlin: Mouton de Gruyter.
- Blasco Ferrer, Eduardo. 1984. *Storia linguistica della Sardegna*. Tübingen: Niemeyer.
- Blasco Ferrer, Eduardo. 1995. Sardisch/Il sardo. In Günter Holtus, Michael Metzeltin & Christian Schmitt (eds.), *Lexikon der romanistischen Linguistik*, vol. 2.2: Die einzelnen romanischen Sprachen und Sprachgebiete vom Mittelalter bis zur Renaissance, 239–271. Tübingen: Niemeyer.
- Blasco Ferrer, Eduardo (ed.). 2003. *Crestomazia sarda dei primi secoli*, vol. 1: Testi – Grammatica Storica – Glossario. Nuoro: Ilisso.
- Cacho Blecua, Juan Manuel (ed.). 2003. *Obra sacada de las crónicas de San Isidoro, de Don Lucas, Obispo de Tuy*. Zaragoza: Universidad de Zaragoza.
- Chomsky, Noam. 2000. Minimalist inquiries: The framework. In David Michaels Martin Roger & Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.
- Cirillo, Robert J. 2009. *The syntax of floating quantifiers: Stranding revisited*. https://pure.uva.nl/ws/files/893235/63916_08.pdf (14 June, 2021).
- CORDE. 2021. *Corpus diacrónico del español*. Madrid: Real Academia Española. <http://corpus.rae.es/cordenet.html> (14 June, 2021).
- Corver, Norbert. 2010. Dressed numerals and the structure of universal numeric quantifiers. In C. Jan-Wouter Zwart & Mark de Vries (eds.), *Structure preserved. Studies in syntax for Jan Koster*, 91–99. Amsterdam: John Benjamins.
- Doetjes, Jenny. 1997. *Quantifiers and selection: On the distribution of quantifying expressions in French, Dutch and English*. Leiden University. (Doctoral dissertation). <https://openaccess.leidenuniv.nl/handle/1887/1973> (14 June, 2021).
- Farina, Luigi. 1973. *Vocabolario Nuorese-Italiano*. Sassari: Gallizi.
- Franco, Irene, Olga Kellert, Guido Mensching & Cecilia Poletto. 2016. On (negative) indefinites in Old Italian. Paper presented at the 38th DGfS conference, February 24–26, 2016. University of Konstanz.
- Franco, Irene & Cecilia Poletto. 2016. Optional negative concord with preverbal n-words. Paper presented at the mini-workshop on negative quantification, June 1, 2016. Goethe University, Frankfurt am Main.
- Giannakidou, Anastasia. 2000. Negative ... concord? *Natural Language and Linguistic Theory* 18. 457–523.
- Giannakidou, Anastasia. 2006. N-words and negative concord. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, 327–392. Malden, MA: Blackwell.

- Giannakidou, Anastasia. 2020. Negative concord and the nature of negative concord items. In Viviane Deprez & M. Teresa Espinal (eds.), *The Oxford handbook of negation*, 458–478. Oxford: Oxford University Press.
- Gianollo, Chiara. 2016. Negation and indefinites in late Latin. *Pallas* 102. 277–286. DOI: 10.4000/pallas.3757.
- Grandgent, Charles Hall. 1907. *An introduction to Vulgar Latin*. Boston: D. C. Heath & Company.
- Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
- Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*. MIT. (Doctoral dissertation).
- Hofmann, Johann Baptist. 1972. *Lateinische Grammatik: Lateinische Syntax und Stilistik*. Neubearbeitet von Anton Szantyr. Verbesselter Neudruck der 1965 erschienenen ersten Auflage, vol. 2. München: C.H. Beck'sche Verlagsbuchhandlung.
- Jonckbloet, W. J. A. 1854. *Guillaume d'Orange, Chansons de geste des XI^e et XII^e siècles*, vol. 1. The Hague: Martinus Nyhoff.
- Jones, Michael Allan. 1993. *Sardinian syntax*. London/New York: Routledge.
- Loi Corvetto, Ines. 1992. Prassi scrittoria nel XIV secolo: Lingua e cultura nel giudicato sardo di Arborea. *Linguistica* 32. 177–196.
- Lupinu, Giovanni (ed.). 2017. *Corpus ATLiSOr: Archivio Testuale della Lingua Sarda delle Origini*. Florence: Istituto Opera del Vocabolario Italiano/Consiglio Nazionale delle Ricerche. <http://atlisorweb.oiv.cnr.it/> (14 June, 2021).
- Malkiel, Yakov. 1948. *Hispanic algu(i)en and related formations: A study of the stratification of the Romance lexicon in the Iberian peninsula*. Berkley, CA: University of California Press.
- Mensching, Guido. 2005. Remarks on specificity and related categories in Sardinian. In Klaus von Heusinger, Georg Kaiser & Elisabeth Stark (eds.), *Nereus II: Specificity and the evolution/emergence of nominal determination systems in Romance*, 81–106. (Proceedings of the international workshop at the Freie Universität Berlin, October 8th–9th, 2004). Konstanz: Universität Konstanz. https://kops.uni-konstanz.de/bitstream/handle/123456789/3798/AP_119.pdf (14 June, 2021).
- Mensching, Guido. 2023. The Old Sardinian quantifier *tot(t)u* 'all': (Lack of) agreement and a $\forall$ NumQ-construction. *Journal of Historical Syntax* 7(16). 1–26. DOI: 10.18148/hs/2023.v7i6-19.154.
- Mensching, Guido & Eva-Maria Remberger. 2016. Sardinian. In Adam Ledgeway & Martin Maiden (eds.), *Oxford guide to the Romance languages (OGR)*, 270–291. Oxford: Oxford University Press.

- Meyer-Lübke, Wilhelm. 1902. Zur Kenntnis des Altlogudoresischen. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften zu Wien, Phil. Hist. Kl.* 145. 1–76.
- Monaci, Ernesto. 1955. *Crestomazia italiana dei primi secoli: Con prospetto grammaticale e glossario*. New edn. Naples: Società editrice D. Alighieri.
- Moseley, Christopher (ed.). 2007. *Encyclopedia of the world's endangered languages*. London: Routledge.
- Norberg, Dag. 1944. *Beiträge zur spätlateinischen Syntax*. Uppsala: Almqvist & Wiksell.
- OVI. 2021. *Opera del Vocabolario Italiano textual database*. <https://artfl-project.uchicago.edu/content/ovi> (1 December, 2021).
- Ramat, Paolo. 1997. Why *veruno* means ‘nobody’. *Romance Philology* 51. 1–14.
- Rindler-Schjerve, Rosita. 1987. *Sprachkontakt auf Sardinien: Soziolinguistische Untersuchung des Sprachenwechsels im ländlichen Bereich*. Tübingen: Niemeyer.
- Rohlf, Gerhard. 1969. *Grammatica storica della lingua italiana e dei suoi dialetti*, vol. 2: *Morfologia*. Translated by T. Francesci. Turin: Einaudi.
- Saba, Agostino. 1927. *Montecassino e la Sardegna Medioevale: Note storiche e codice diplomatico Sardo-Cassinese*. Montecassino: Badia di Montecassino.
- Sanna, Antonio. 1975. *Il dialetto di Sassari (e altri saggi)*. Cagliari: Trois.
- Sophocles, Evangelinus Apostolides. 1900. *Greek lexicon of the Roman and Byzantine periods (from B.C. 146 to A.D. 1100)*. Memorial Edition. New York: C. Scribner's Sons.
- Szabolcsi, Anna. 1997. Strategies for scope taking. In Anna Szabolcsi (ed.), *Ways of scope taking*, 109–154. Dordrecht: Foris.
- Tola, Pasquale. 1850. *Codice degli Statuti della Repubblica di Sassari*. Cagliari: Tipografia di A. Timon.
- Virdis, Maurizio (ed.). 2003. *Il Condaghe di Santa Maria di Bonarcado*. Nuoro: Ilisso.
- Wagner, Max Leopold. 1907. *Lautlehre der südsardischen Mundarten: Mit besonderer Berücksichtigung der um den Gennargentu gesprochenen Varietäten*. Halle: Niemeyer.
- Wagner, Max Leopold. 1938. Flessione nominale e verbale del sardo antico e moderno. *Italia dialettale* 14;15. 93–70, 207–247.
- Wagner, Max Leopold. 1997. *La lingua sarda: Storia, spirito e forma, nuova edizione a cura di Giulio Paulis*. Nuoro: Ilisso.
- Wolfe, Sam. 2015. The Old Sardinian *condaghes*: A syntactic study. *Transactions of the Philological Society* 113(2). 177–205. DOI: 10.1111/1467-968X.12046.

Chapter 7

The Brazilian Portuguese present perfect: From nominal to verbal pluractionality

Malte Rosemeyer

Freie Universität Berlin

Martin Becker

University of Cologne

We analyze the semantic change undergone by the present perfect in Brazilian Portuguese (BP). While the present perfect in earlier stages of Portuguese is usually described as a resultative perfect, the construction has acquired iterative and durative readings in Modern Portuguese. We analyze the distribution of more than $n = 850$ occurrences of the present perfect in a diachronic corpus of BP theater texts, ranging from the 19th until the 21st century. We develop a bottom-up approach towards calculating the likelihood for a given context of the present perfect to express pluractional readings in the nominal and verbal domains. We measure the correlations between this parameter and time, as well as register. In doing so, we are able to establish a more precise model of the semantic change experienced by the present perfect in BP. The results from the analysis provide empirical evidence for Amaral & Howe's (2012) claim that the reanalysis of the present perfect involved a transfer from nominal to verbal pluractionality. Additionally, our analysis reveals that register had an important influence on this change. This finding leads us to hypothesize that the change towards pluractional readings in the BP present perfect was facilitated by intensive contact with European Portuguese during the second half of the 19th century, i.e. the so-called period of the *Império do Brasil* ('Empire of Brazil').



1 Introduction

In earlier stages of Portuguese, the present perfect (henceforth PPC, for the Portuguese term *pretérito perfeito composto*), was used as a resultative perfect. For instance, in (1), the PPC expresses the present result of an event that has occurred exactly once in the past, and which was concluded a while before speech time.

- (1) Tycho Brahe Parsed Corpus of Historical Portuguese, 16th c., Amaral & Howe (2012: 28)

Eu ey por bem que Nicolao Jusarte, fidalgo de minha
I have.PRS.IND.1SG for good that Nicolao Jusarte, nobleman of my.F
casa, a que tenho feito merce da capitania
house, to who have.IND.PRS.1SG do.PTCP grace of.DET.DEF.F.SG captaincy
de um dos navios que vão pera
of DET.DEF.M.SG of.DET.DEF.M.PL ships that go.IND.PRS.3PL to
a India nesta armada d'outubro, vaa
DET.DEF.F.SG India in.DEM.F.SG fleet of.october go.IND.PRS.3SG
no navio do Porto
in.DET.DEF.M.SG ship of.DET.DEF.M.SG Porto
'I order that Nicolau Jusarte, nobleman of my house, whom I have
awarded the honor of being the captain of one of the ships that will go to
India in the October fleet, shall go on the ship from Porto'

Examples such as (1) are ungrammatical in Modern Portuguese due to a historical change in the usage contexts of the PPC from resultative to durative or iterative readings. Consider, for instance, example (2), taken from the translation of an interview with designer Jony Ive, where the original English sentence is inflected for present perfect progressive (*we've been doing*), whereas the Portuguese translation uses the PPC. The use of the present perfect progressive in the original text suggests that the speaker is talking about an event whose beginning lies in the past (Ive's work with Apple) and has been continuing until speech time.

- (2) Interview with Jony Ive, *Corpus do Português* (Davies 2016–, original English text from <https://www.ft.com/content/0b20032e-98cf-11e9-8cfb-30c211dcd229>, accessed 20 September, 2020)

Há algumas áreas que são paixões
 have.IND.PRS.3SG DET.INDF.M.PL areas that be.IND.PRS.3PL passions
 naturais para mim. O trabalho que tenho feito
 natural for me. DET.DEF.M.SG work that have.IND.PRS.1SG do.PTCP

com a tecnologia de wearables – com a
 with DET.DEF.F.SG technology of wearables – with DET.DEF.F.SG
 tecnologia a tornar-se mais pessoal, é
 technology to become.INF-REFL.3 more personal be.IND.PRS.3SG
 uma inevitabilidade que se torne usada
 DET.INDF.M.PL inevitability that REFL.3 become.SBJ.PRS.3SG use-PTCP.F
 ‘There are some areas that are personal natural passions for me. The
 work that we’ve been doing with wearable technology – with technology
 becoming more personal, there is an inevitability that it becomes worn’
 [=original English text]’

The change from resultative to durative or iterative readings has received great attention in the linguistic literature. In a recent proposal, Amaral & Howe (2012) analyze the development of the PPC in terms of the notion of “pluractionality”, understood as the potential for a sentence to express multiple events. They claim that the PPC acquired pluractional readings in contexts that are structurally ambiguous; in particular transitive contexts with a direct object that is inflected for masculine and singular, but can still be interpreted as expressing plural. For instance, in their example (3) below, the direct object *hospital* is modified by the universal quantifier *cada* ‘each’. As a result, the PPC *tenho provido* in this example can be understood as either expressing a resultative reading (‘each hospital has a physician now’) or a pluractional reading (‘I have repeatedly granted the hospitals a physician’).

- (3) Tycho Brahe Parsed Corpus of Historical Portuguese, 16th c. (Amaral & Howe 2012: 43)

tenho **provido** cada hospital de seu físico, que
 have.IND.PRS.1SG grant.PTCP each hospital of POSS.M.SG physician, who
 são os abades, retores, vigários e curas
 be.IND.PRS.PL DEF.DET.M.PL abbots, rectors, vicars and priests
 ‘I have granted each hospital with a physician, who are the abbots,
 rectors, vicars, and priests’

Amaral & Howe (2012) assume contexts such as (3) to be pivotal to the semantic change undergone by the PPC, in that hearers of such examples could have reanalyzed the PPC as expressing pluractionality in these contexts. This would involve a “transfer of semantic plurality in the nominal domain to the verbal one” (Amaral & Howe 2012: 51), in that an interpretive property formerly associated with an argument in the sentence would now come to be conventionally associated with the use of the PPC.

Crucially, however, Amaral & Howe (2012) offer little quantitative evidence for the relevance of this mechanism and, indeed, do not map out the semantic change itself in diachrony. The present paper establishes a data-driven description of the historical change undergone by the PPC in Brazilian Portuguese. In doing so, it aims to test Amaral & Howe's (2012) hypothesis of a transfer from nominal to verbal pluractionality. In particular, this hypothesis makes the prediction that in earlier stages of the change, the PPC occurs in contexts that denote nominal pluractionality (i.e. transitives with direct objects that allow for a plural reading) with sufficient frequency as to enable a reanalysis that would then affect the usage of the PPC in all contexts. A second prediction would be that whereas the likelihood of the PPC to be used in nominal pluractionality contexts did not increase over time, its likelihood to be used in verbal pluractionality contexts (corresponding to the eventual conventional meaning of the PPC) did increase over time.

We investigate these predictions in a large corpus of BP theater plays, which approximate the spoken language of the respective periods, between the 19th and 21st century. On the basis of a careful qualitative analysis of the use of the PPC in these plays, we establish a bottom-up quantitative operationalization of the likelihood for a given context to express nominal or verbal pluractionality. Our results confirm the predictions made by Amaral & Howe's (2012) hypothesis, according to which in earlier texts, the PPC is more likely to be used in contexts associated with nominal pluractionality than in later texts. At the same time, we find a general increase in the likelihood of the PPC to be used in contexts associated with verbal pluractionality.

Our corpus of BP theater plays also allows us to measure the sociolinguistic environment of the semantic change undergone by the PPC. This analysis allows for an explanation of a curious inconsistency in the development of the PPC towards usage in verbal pluractionality contexts, namely a notable decrease in speed of the change in texts from the first half of the 20th century. In particular, we measure the degree to which each theater play represents conceptual orality, understood as the degree to which the scripted language in the plays is presented as highly planned and edited (Koch & Oesterreicher 1985). A high degree of conceptual orality indicates a low adherence to the linguistic norm of written language at that time and, in contrast, a more authentic representation of actual spoken language. By integrating this measurement into ordinal logistic regression models, we are able to establish that the semantic change towards pluractionality proceeded in two qualitatively different stages: in a first stage (1850–1890), we find a strong increase of the use of the PPC in verbal pluractionality contexts only in texts with low conceptual orality. In contrast, in a second stage

(1950–2016) of expansion towards verbal pluractionality contexts, this process is implemented to a stronger degree in texts with high conceptual orality. We interpret this finding as indicative of an influence of the stronger contact between Brazilian and European Portuguese during the 19th century, i.e. the so-called period of the *Império do Brasil* ('Empire of Brazil'). Given that the PPC experienced the change towards expression of verbal pluractionality earlier in European than in Brazilian Portuguese, we hypothesize that the stronger contact with the prestigious European norm during the *Império do Brasil* caused conservative authors to adopt usage patterns of the PPC associated with the European norm.

2 Previous studies

The Portuguese PPC is a well-described construction, which has been analyzed in a number of studies. Both grammars (Hundertmark-Santos Martins 1982: 180, Gärtner 1998: 23, Perini 2002: 249–250) and linguistic studies (Boléo 1936: 127, Ilari 2001a: 66, Novaes & Nespoli 2014, Santos 2008) establish that, in Modern Portuguese, the PPC can express either a durative (4a) or a non-durative reading, such as the iterative reading in (4b).¹

- (4) a. Durative reading (Hundertmark-Santos Martins 1982: 180)
Tem estado muito calor.
 have.IND.PRS.3SG be.PTCP much heat
 'Recently, it has been very hot.'
- b. Iterative reading (<http://caras.sapo.pt/famosos/2011-07-04-claudia-vieira-tenho-vivido-bons-momentos-desde-que-a-maria-nasceu>, last access 4 November 2019)
Tenho vivido bons momentos desde que a
 have.IND.PRS.1SG live.PTCP good moments after that DET.DEF.F.3SG
 Maria nasceu.
 Maria be.born.IND-PST.PFV.3SG
 'I have experienced many good moments since Maria's birth.'

These prototypical readings are described as manifestations of an abstract semantics of the PPC that reunites several characteristics that have been shown to be relevant in the literature (Ilari 1999, 2001a,b, Barbosa 2008, Schmitt 2010, Molsing 2010, Cabredo Hofherr et al. 2010, Barbosa 2012, Amaral & Howe 2012, Oliveira & Leal 2012, Olbertz 2018, Becker 2020):

¹There are a number of non-durative readings that the PPC can obtain. The qualitative analysis in §4 will describe these readings in more detail, using examples from our historical corpus.

- The PPC expresses a plurality of events (or sub-events) of the type described by the predicate because it introduces an operator expressing indeterminate quantification of events (or sub-events) (see Barbosa 2008: 98, Barbosa 2012: 182, Becker 2020).
- The PPC introduces a temporal interval that extends from an initial point in the past (left boundary) to speech time (right boundary) and can, in the right context, even be interpreted as generating the implicature that the event extends until the future (Becker 2017: 32, Olbertz 2018: 489). This interval is asserted, equaling the so-called “perfect time span” (see, for instance, von Stechow 1999).
- The different readings depend on the aspectual properties of the verbal predicates, in that, for instance, without further modification stative predicates generate durative readings, as in (4a).

Recent analyses make use of the notion of pluractionality, which was introduced into semantics by Lasnik (1995). Pluractionality can be described as an umbrella term for various kinds of readings that involve event plurality, such as iterative and habitual readings. There is no consensus regarding the exact definition of pluractionality. While Cabredo Hofherr et al. (2010: 83–84) define pluractionality as an indeterminate plurality of events, thus excluding contexts with definite quantification such as cardinal quantification of the type ‘two times’, other authors such as Bertinetto & Lenci (2012) define the term more broadly. In their words, pluractionality indicates that “the same event repeats itself in a number of different situations”, not excluding definite quantification. In this paper, we will use this broader definition of pluractionality, and broaden it even further to contexts in which the pluralizing reading derives from a pragmatic inference. This inference has sometimes been described as a coercion mechanism (Pustejovsky 1995, Michaelis 2004) in that pluractional readings are generated on the basis of the interaction between certain verb forms, predicate types and adverbial expressions (Laca 2006). The semantic change of the PPC from resultative to pluractional readings, which stands in the center of interest of this paper, has likewise received some attention in the recent literature (Barbosa 2014, Amaral & Howe 2012, Becker 2017, Olbertz 2018: 486–493). These studies differ as regards the variety of Portuguese that is being analyzed, the time frame, methodology (qualitative vs. quantitative), and research questions.

Whereas Amaral & Howe (2012), Becker (2017), Olbertz (2018), and Becker (2020) are mostly interested in motivating the semantic change from resultative

to pluractional readings, Barbosa (2014) aims at establishing a complete description of the changes undergone by the PPC in Brazilian Portuguese between the 16th and 20th century.² This is done by opposing the PPC and the simple past tense (the “pretérito perfeito simples”, or PPS), using variationist methodology. First, her results demonstrate that in relation to the PPS, the usage frequency of the PPC is higher in earlier stages of BP than after the 20th century. Between the 18th and 20th century, the usage frequency of the PPC relative to the PPS drops from 14 percent to 5 percent. Perhaps due to the fact that texts from the 17th and 18th century do not really reflect a Brazilian Portuguese norm (see footnote 2), her results for these centuries display a somewhat mixed tendency.

Second, Barbosa (2014: 93) analyzes the change in the semantics of the PPC. Her results demonstrate the expected change from non-pluractional towards pluractional readings; whereas she analyzes 51 percent of the occurrences of the PPC in the 16th century as expressing a (resultative) perfective reading, she documents a steady decrease of the usage frequency of the PPC in these contexts until the 20th century, in which she does not find any tokens of the PPC expressing this reading. In contrast, she finds a mild increase in the relative frequency of iterative readings (which are found to be frequent already in early texts) and a strong increase in the relative frequency of durative readings (from 7 percent in the 16th century to 40 percent in the 20th century).

While these numbers are very convincing, it has to be noted that they represent the author’s interpretation of the data and can consequently be considered as highly subjective. This is problematic because (a) sentences involving the PPC are frequently ambiguous between the different readings, a fact that is actually at the heart of the semantic reanalysis of the PPC (see the discussion of Becker 2017 and Amaral & Howe 2012 below) and (b) there is no consensus in the literature with respect to the precise range of functions expressed by the PPC. For instance, Olbertz (2018: 487) claims that “[i]t is only in the 20th-century data that we have unambiguous cases of the iterative (or continuous) perfect, in which the event is implied to persist”. Third, Barbosa (2014) also analyzes the influence of the degree of formality of the texts on the opposition between the PPC and the PPS, although her basis for the classification of the corpus into texts that are [+formal] or [–formal] is unclear. She claims that formality influenced the opposition only in the 16th century and that, after the 17th century, formality did not have an influence on the use of the PPC.

² As noted by Barbosa (2014: 85, fn 1) herself, the label “Brazilian Portuguese” is somewhat misleading for texts written in the 16th and 17th century because at that time, a Brazilian Portuguese norm did not exist. Her “Brazilian Portuguese” texts from this period are texts written by European Portuguese authors living in Brazil.

Becker (2017) provides a fine-grained qualitative analysis of the semantics of the PPC, distinguishing several readings. He describes in greater detail uses of the PPC that he calls “summative”, in which the pluractional reading derives from an inference on the basis of the combination of adverbial expressions and sentence arguments. Consider Becker’s example (5), taken from the corpus *do português* (Davies 2016–). Here, the author uses the PPC to describe various occurrences of the action ‘making a will’ that are situated in a time interval that ranges from some moment in the past until before the speech act (*athé aqui* ‘up until here’). Note that the PPC in (5) indicates indefinite quantification, in that the exact number of testaments is unknown (and, indeed, irrelevant here). The pluractional reading mostly derives from the plurality of the direct object constituent, marked also by agreement on the participle *feitos*. According to Becker, such uses of the PPC are extremely rare in the 16th century and can be considered innovative.

- (5) *Corpus do português*, 16th c. (Becker 2017: 29)
- | | | | | | | |
|------|---------------------|------------------|--------|-------------|------|-------|
| e | ruogo | todollos | outros | testamentos | que | athe |
| and | ask.for.IND.PRS.1SG | all.DET.DEF.M.PL | other | wills | that | until |
| aqui | hey | feitos | | | | |
| here | have.IND.PRS.1SG | do.PTCP.M.PL | | | | |
- ‘and I ask for all of the other wills that I have made until today’

Crucially, Becker (2017: 33) observes an increase in the usage frequency of such summative readings of the PPC in the 17th century and claims that these readings can be considered the starting point of the semantic change of the PPC towards the expression of pluractionality. In particular, he observes that durative readings were developed only after the use of the PPC was firmly entrenched in summative contexts.

While using different terminology, Amaral & Howe (2012) model the semantic reanalysis of the PPC in a very similar manner. In particular, they identify contexts such as example (3), repeated below as (6), as “onset” contexts for the change from resultative to pluractional readings. Example (6) seems to be a clear instance of the contexts called “summative” by Becker (2017); in particular, the pluractionality value of the PPC derives from a pragmatic inference on the basis of the implicit plurality of the direct object *cada hospital*. Like (5), example (6) is ambiguous between this innovative pluractional reading and the original resultative reading, which is a hallmark of “bridging contexts”, i.e. contexts that are characterized as intermediate steps in processes of semantic change (Heine 2002).

- (6) TBCHP, 16th c. (Amaral & Howe 2012: 43)

tenho **provido** cada hospital de seu físico, que
 have.IND.PRS.1SG grant.PTCP each hospital of POSS.M.3SG physician who
 são os abades, retores, vigários e curas
 be.IND.PRS.PL the abbots rectors vicars and priests
 ‘I have granted each hospital with a physician, who are the abbots,
 rectors, vicars, and priests’

Amaral & Howe (2012: 40–48) provide a testable hypothesis as to the nature of this semantic reanalysis. Crucially, the resultative and pluractional readings of example (6) presuppose the same kind of complex event structure, namely a process by which every hospital was assigned a physician. The difference between the two interpretations resides in the fact that the pluractional reading asserts that this assignment process was carried out multiple times with different referents (i.e., physician1 was assigned to hospital3, physician2 to hospital4, etc.). In contrast, the resultative reading merely asserts the result of these processes, i.e. that each hospital has been assigned a physician. Since in the resultative reading, the truth of the existence of these sub-events is not evaluated, it is perfectly compatible with both a single-event reading (all physicians were assigned in the same event) and a plural-event reading (each doctor was assigned separately). According to Amaral & Howe (2012: 43–43), the change from a resultative to a pluractional reading can then easily take place in contexts in which the plural event interpretation is more likely for pragmatic reasons. For instance, in example (6), it seems unlikely that the subject referent was able to assign all physicians at the same time.

The idea that contexts of the type instantiated in (6) constitute bridging or, in Amaral & Howe’s (2012) terms, “onset” contexts entails that the semantic change experienced by the PPC can also be described as change from this type of pluractional reading to another type of pluractional reading, i.e. contexts that are no longer compatible with a resultative reading. As an example for such a “switch context” (Heine 2002: 85), consider (7) taken from Becker’s (2017) study. Both instances of the PPC in (7) are incompatible with a resultative reading because the predicates *pensar* and *estar* do not entail resultant states. They also differ from example (6) in that the plurality of the event is not indicated by a plurality of one of the arguments of the verb.

- (7) *Corpus do português*, 20th c. (Becker 2017: 37)
- Esta é uma questão que **tenho**
DEM.F.SG be.IND.PRS.3SG DET.INDF.F.SG question that have.IND.PRS.1SG
pensado muito ao longo de 7 anos que
think.PTCP much at.DET.DEF.M.SG long of seven years that
tengo estado na Microsoft
have.IND.PRS.1SG be.PTCP in.DET.DEF.F.SG Microsoft
'That's a question that I have been thinking about a lot during the seven
years that I have been at Microsoft'

Consequently, the semantic change leading from examples such as (6) to examples such as (7) can be described as a conventionalization of the pluractional reading of the PPC. In other words, the pluractional reading of the PPC became independent from contextual cues such as plurality of the argument(s) in the sentence and has thus become part of the conventional meaning of the PPC. This change crucially involves a "transfer of semantic plurality in the nominal domain to the verbal one" (Amaral & Howe 2012: 51), given that the trigger of the pluractional reading no longer resides in the nominal arguments in the sentence, but rather in the use of the PPC itself.

To summarize, studies on the development of the PPC assume that the semantic change undergone by the PPC can be interpreted in terms of a semantic reanalysis enabled by bridging contexts in which the pluractional reading derives via pragmatic inference from the (semantic) plurality of one or more of the arguments of the verb. The change from such "summative" to purely iterative and durative readings crucially involved a transfer from nominal to verbal plurality.

It is this hypothesis, namely the transfer from nominal to verbal plurality, that will be at the heart of the analysis presented in this paper. In particular, the assumption that contexts in which the pluractional reading derived via pragmatic inference from the presence of (semantically) plural nominal arguments served as bridging contexts for the conventionalization of the pluractional reading, makes the specific predictions that (a) the PPC is more likely to occur in contexts characterized by nominal pluractional readings in earlier texts than in later texts, and (b) likelihood for the PPC to be used in contexts associated with verbal pluractionality is expected to increase over time. Note that none of the papers discussed until now have given quantitative empirical evidence that might confirm these predictions.

As is well known in historical linguistics, the sole mapping out of a reanalysis process does rarely do justice to the complexities of the description of the entire process of historical change, which is why recent approaches call for longitudinal descriptions that also analyze sociohistorical parameters. In particular,

grammaticalization processes are non-teleological in the sense that they have more than one possible outcome (Collins 2019). As a result, it is extremely important to consider both the source constructions in such processes (Cristofaro 2019) and map out the evolution of these processes over time, which can be affected by phenomena such as language contact (Thomason 2010, Poplack et al. 2011, van der Auwera & Gast 2012) or socio-stylistic attitudes involved in processes such as latinization (Cornillie & Drinka 2019). Consequently, we were also interested in the influence of one parameter, the degree of conceptual orality of the texts, on the development of the semantic change of the PPC, with the working hypothesis that the semantic change is expected to be implemented to a greater degree in texts that display a high degree of conceptual orality and consequently, a low degree of formality (see Rosemeyer 2019a for discussion).³

3 Data and periodization

In order to provide quantitative evidence for the assumption that the semantic change of the PPC towards pluractionality crucially involved a transfer of pluractional readings in nominal contexts to verbal contexts, we extracted all tokens of the PPC from two diachronic corpora of Portuguese: the PorThea, a corpus of historical theater plays (Rosemeyer 2019b), and the Genre/Historical section of the Corpus do português, a multi-genre corpus with historical depth (Davies 2016–, henceforth CdP). Table 1 below gives an overview of the two corpora.

Table 1: Overview of the PorThea and CdP corpora

	PorThea	CdP
Corpus size	~3.3 million words	~45 million words
Varieties of Portuguese	EP, BP	EP, BP
Genre	Theater plays	Different genres, including spoken language
Time depth	18 th –21 st c.	13 th –20 th c.
Automatic annotation	None	PoS tagging, lemmatization

³Recall that the analysis regarding this parameter proposed in Barbosa (2014) did not yield conclusive results.

As evident in Table 1, the PorThea corpus is considerably smaller and less-well annotated than the CdP. Its main advantage, however, is its thematic consistency: given that theater plays approximate orality, a corpus of theater plays can be expected to give a better indication of patterns of change in the spoken language at that time than other written sources (Kytö 2011: 432). When using data from the CdP, it is possible that observed changes in usage frequency and distributional patterns are actually due to changes in the textual substrate, making such changes “apparent changes” (see Szmrecsanyi 2016, Rosemeyer 2019a). For this reason, we used the data from the PorThea for the main analyses conducted in this paper and the CdP data merely for sanity checks.

Extraction of all tokens of the PPC from the Brazilian section PorThea led to a total result of $n = 857$ tokens using regular expressions. Figure 1 visualizes the historical development of the usage frequency of the PPC in these data.

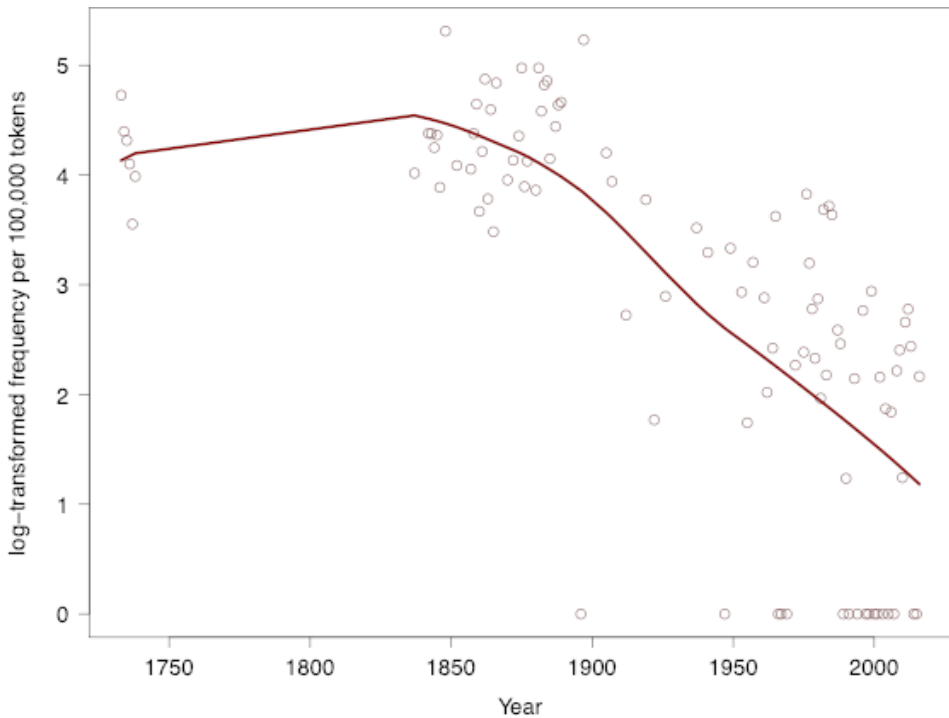


Figure 1: Historical development of the log-transformed usage frequency per 100,000 tokens of the PPC in the Brazilian section of the PorThea corpus. Points represent mean usage frequencies per year, whereas the line represents the result from a local polynomial regression analysis summarizing the trend.

It is important to note that the PorThea does not contain texts for the period between 1750 and about 1830, which is why no data points are given in the plot for this period.⁴ However, if we compare the mean usage frequencies of the PPC in plays dating from the first half of the 18th century and the mid-19th century, we find a weak increase. In contrast, the analysis suggests a strong and relatively linear decrease in the usage frequency of the PPC from the second half of the 19th century to Present-Day BP. This finding, which reaches statistical significance,⁵ thus corroborates the previous findings by Barbosa (2014) as regards the decreasing usage frequency of the PPC in BP.

Although usage frequency and grammatical productivity of a construction typically stand in a relationship to each other, a higher usage frequency does not always indicate higher grammatical productivity (Barðdal 2008). For instance, grammaticalization processes are usually assumed to involve a rise in usage frequency because the semantic change leads to an expansion of the use of the construction into new usage contexts (so-called “host class expansion”, Himmelmann 2004). In our case, such a “host-class expansion” could be understood as an expansion of PPC usage to more verb classes. However, it has also been shown that in incipient grammaticalization processes, prefabs, i.e. conventionalized form-function pairings which would usually be described as rather unproductive, lead the way (Bybee & Torres Cacoullos 2009). Likewise, lexicalization processes can involve an increase in usage frequency that does not reflect an increase in productivity. As a result, a complete description of any linguistic change needs to not only rely on usage frequencies, but also on productivity measures.

One such productivity measure, which is common in corpus linguistics, is type-token ratio (TTR). The type-token ratio is calculated by dividing the total number of types (words, constructions etc.) by the total number of tokens (McEnery & Hardie 2012: 253), yielding a measure between 0 and 1. When analyzing verbal constructions, TTR can be a measure of productivity in terms of the degree to which that construction can be used with all verbs in a language. To achieve this, we divided the number of verb lemmas found in the PPC in each year of the corpus by the total number of the PPC found in that year. Whereas a higher TTR indicates a high productivity (wider range of verbs found in the PPC), a lower TTR indicates a lower productivity (lower range of verbs found in the PPC).

⁴The corpus size in words by century is: $n_{18^{th}} = 175,891$, $n_{19^{th}} = 787,015$, $n_{20^{th}} = 747,110$, $n_{21^{st}} = 948,485$. Raw usage frequencies of the PPC per century are: $n_{18^{th}} = 106$, $n_{19^{th}} = 570$, $n_{20^{th}} = 116$, $n_{21^{st}} = 65$.

⁵Correlation testing was done using Kendall’s τ because both variables were not normally distributed (cf. Gries 2009). There is a significant negative correlation between the date of the text and the log-transformed normalized frequency of the PPP per 100,000 words (Kendall’s $\tau = -0.371$, $z = -5.412$, $p_{\text{one-tailed}} < 0.001$).

Figure 2 visualizes the development of the TTR of the PPC by year in our corpus. It demonstrates an increase in TTR between 1850 and 1950, followed by a steep decrease after 1950.⁶

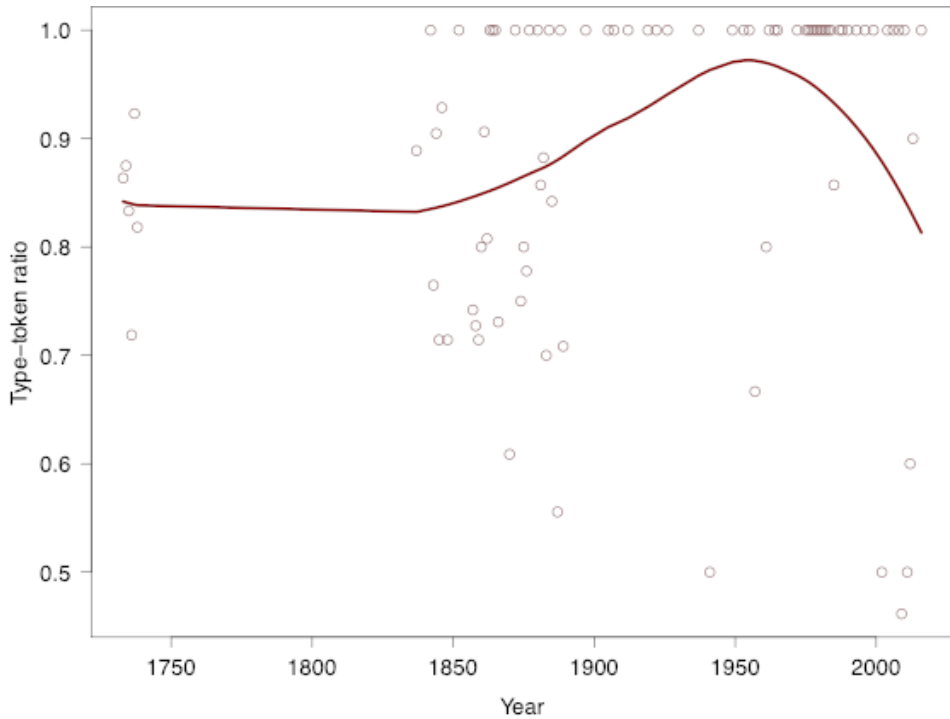


Figure 2: Historical development of the type-token ratio (TTR) of the PPC in the Brazilian section of the PorThea corpus. Points represent TTR ratios per year, whereas the line represents the result from a local polynomial regression analysis summarizing the trend.

If we compare the development of the TTR and the usage frequency of the PPC (see Figure 1 above), we see that these two changes are not always correlated. Thus, between 1850–1950, we find that as the usage frequency of the PPC decreases, its TTR increases. In contrast, after 1950, the PPC decreases both in usage frequency and TTR. This finding suggests that the change by which the

⁶Again, there is a significant correlation between TTR and year of publication of the play. In order to capture the non-linearity of the trend visualized in Figure 2, we calculated a linear regression model predicting TTR by year of publication. Given the distribution of TTR found in Figure 2, the variable YEAR was modeled as a third-degree polynomial. A significant effect was found ($F(3, 80) = 3.328, p < 0.05$), with an R^2 of 0.111.

PPC decreased in usage frequency after 1850 might actually consist of two qualitatively different, subsequent, changes.

As the visualizations in Figures 1 and 2 have shown, the Brazilian section of the PorThea corpus does not contain data for the time period between about 1750 and 1830, which is a problem for more fine-grained analyses of the distribution of the PPC. We consequently decided to eliminate all 18th century tokens of the PPC from the subsequent quantitative analyses in §5 and §6, leading to a new total of $n = 751$ tokens.

The data from the CdP confirm the general trend towards a lower usage frequency of the PPC. Thus, the PPC reaches 530 words per million in the 18th century CdP, 461 words per million in the 19th century, and 356 words per million in the 20th century. The historical data from the CdP do not distinguish between European and Brazilian Portuguese. However, it is possible to extract comparative frequency measures for the 20th century; here the CdP data show the PPC to be much more frequent in EP than in BP (450 words per million in EP, 265 words per million in BP). This seems to suggest a greater productivity of the PPC in EP than in BP.

4 Pluractional readings and (some of) its contexts

In this section, we try and establish the contextual parameters in the nominal and verbal domains that indicate a pluractional reading and identify some typical contexts and constellations of these pluractional readings arising. In many cases, several linguistic clues interact in producing a reading of plurality. All of the examples cited in this section are from our PorThea data.

In the nominal domain, a pluractional reading is inferred from the plurality of one or more of the arguments in the sentence (subject, direct object, indirect object). A very important role is played by the structure and semantics of the subject. A typical context combines a plural subject NP with a reflexive verb. In the example (8), the interaction of the plural subject with the reflexive verb produces a distributive reading. The subject NP *nós todos* is interpreted as referring to every single individual (corresponding to *cada um de nós*), which is predicated to be involved individually in a sub-event as part of an overarching collective commitment event. Consequently, the collective commitment event implies an indeterminate number or set of sub-events with its individual participants. This configuration corresponds to the structure in (9), which ascribes a commitment-sub-event to each member of the quantitatively undetermined we-group.

- (8) Martins Pena, *O noviço*, 1845

Nós todos nos temos empenhado
 we all us.REFL.3PL have.IND.1PL make.effort.PTCP
 ‘We have all made an effort’

- (9) {(individual 1| commitment-subevent1), (individual
 2|commitment-subevent2), (..|..), (individual n| commitment- subevent n)}.

The same – though more subtle – effect is produced by the interplay with collective subject-NPs, as in example (10). Here, the NP *o povo* refers to a collective entity which is composed of individuals. At the same time, the temporal adverbial *de algum tempo para cá* marks a typical perfect interval ranging from some moment in the past (left boundary) to the speech time (*para cá*, i.e. the right boundary). It can be inferred from this constellation that the set of members referred to by the collective *povo* changes over time as to its composition so that at each moment different members are involved in the macro-event of SHOWING_DEMOCRATIC_TENDENCIES. In other words, we can attribute different sets of individuals to each time point.

- (10) Artur Azevedo, *A princesa dos cajueiros*, 1880

o povo tem mostrado de algum tempo para
 DET.DEF.M.PL people have.IND.PRS.3SG show.PTCP from some time to
 cá certas tendências democráticas
 here certain tendencies democratic
 ‘from time to time the people have shown certain democratic tendencies’

A plural(ity) reading can also be triggered by a plural direct object NP. In example (11), the speaker not only refers to several acts of favors, but also enhances the quality of the acts (*obséquios imensos*). This is a means of strengthening the inference that the speaker has been granted several favors, each of which required a certain amount of time.

- (11) José de Alencar, *O crédito*, 1857

não poderei pagar-lhe a amizade e os
 not can.IND.FUT.1SG pay-you DET.DEF.F.SG friendship and DET.DEF.M.PL
 obséquios imensos que **nos têm sido feitos**
 presents immense that to.us have.IND.PRS.3SG be.PTCP make.PTCP.M.PL
 ‘I cannot make up to you the friendship and the limitless presents that we have been given’

The direct object is not necessarily a plural noun but may also turn out to be a mass noun, as in example (12). What is important in these contexts, however, is the presence of a quantifying expression, i.e. of a quantifier phrase embedded in the NP. The quantifying expression implies that the macro-event in question falls into several stages in which a part or portion of the direct object is affected by the transitive process in question. In (12), the whole quantity of N (*tôda a prata*) is affected by the CARRY-IN-event (*levar para dentro*). Our available world knowledge suggests that the whole event is structured by sub-events in which portions of the direct object referent partake. Bertinetto & Civardi (2015) introduce a measure δ , which indicates to what degree an incremental theme is affected by a telic event. In our example *levar toda a prata*, the theme, as made explicit by the quantifier, is completely affected (so $\delta = 1$). However, the completion process requires subsequent phases until its culmination point is reached (and the whole quantity of silver has been brought in).

- (12) Martins Pena, *O usurário*, 1846
 Enquanto assim falam, os dois **têm**
 meanwhile so talk.IND.PRS.3PL DET.DEF.M.PL two have.IND.PRS.3PL
levado **tôda a** **prata** para dentro.
 bring.PTCP all DET.DEF.F.SG silver to inside
 ‘While chatting, the two of them have brought all of the silver inside.’

It goes without saying that plurality effects can also be obtained by the indirect object. In example (13), the collective noun *muita gente* refers to a set composed of individuals that are affected individually by the same kind of event.

- (13) Artur Azevedo, *A capital federal*, 1897
 Volte, seu Figueiredo, volte, se não
 turn.back.SBJ.PRS.3SG, dear Figueiredo turn.back.SBJ.PRS.3SG if not
 quer que lhe aconteça o mesmo
 want.IND.PRS.3SG that to.you happen.IND.PRS.3SG DET.DEF.M.SG same
 que me sucedeu e **tem** **sucedido** a
 that to.me happen.IND.PST.PFV.3SG and have.IND.PRS.3SG happen.PTCP to
muita gente!
 many people
 ‘Turn back, dear Figueiredo, turn back, if you do not want that which has happened to me and to many people to also happen to you!’

Another comparable case with an incremental theme is the one in (14). The quantifiers used in this example, namely *um cento de* and *outro tanto* (*um cento de*

cartas, outros tantos pedidos), are very frequent in PPC contexts. Neither *um cento de*, with the meaning ‘approximately one hundred’, nor (*outro*) *tanto* (‘another x instances of N’) specify the precise number of instances of the incremental theme in question. This boils down to the fact that the number of instances of the writing-events (of letters and of requests) remains indeterminate. If the speaker had wanted to indicate a precise quantification of the instances (e.g. exactly 100 letters), he would have resorted to the simple preterit (*Escrevi cem cartas e outro tanto pedidos* ‘I wrote a hundred letters and pleas’).

- (14) Qorpo Santo, *O marinheiro escritor*, 1866

Tenho **escrito** **um** **cento** **de cartas, feito**
have.IND.PRS.1SG write.PTCP DET.INDF.M.SG hundred of letters, do.PTCP
outros tantos **pedidos a** **peessoas que para lá** **vão;** **já**
other so.many pleas to persons that to there go.IND.PRS.3PL already
mandei **de propósito uma** **para tal fim, e nada**
send.PST.PFV.1SG of purpose DET.INDF.F.SG for such end and nothing
tenho **podido** **consequir**
have.IND.PRS.1SG be.able.PTCP achieve.INF
‘I have written some hundred letters, made pleas to people who are going there; I already sent one there on purpose to achieve this aim, and I haven’t been able to accomplish anything’

A particular pluractional setting may be produced in contexts with internal temporal modification, such as a negation operator (see 15). Typically, the negation operator interacts with an explicit temporal expression. In this constellation, the negation operator is in the scope of an all-quantification inferable from the temporal expression. In (15), the speaker asserts that, for all time indexes *i* of an interval *I*, there is no event *e* of the given type, such that *p* holds, i.e. for all *i* ∈ *I*: ¬ *p*(*i*). Therefore, this configuration yields the durative reading that the same state of affairs non-*p* holds during the whole relevant interval set by the expression *até agora*. Once again, we are dealing with a perfect time span ranging from some point in the past to the utterance time of the speaker.

- (15) Martins Pena, *Os irmãos das almas*, 1844

Até agora não tenho **sido** **homem, mas era**
until now not have.IND.PRS.1SG be.PTCP human but be.PST.IPFV.3SG
preciso **sê-lo**
necessary be.INF-it
‘Until now I haven’t been human, but it was important to be it [=a dog]’

A very similar effect can derive from the use of the direct object *nada* ‘nothing’ (see 16). As in example (15), the speaker asserts that a certain event did not happen within a certain time interval, a reading that seems to be highlighted by the resumptive preposing (Leonetti 2017: 908–911) of *nada*.

- (16) Qorpo Santo, *O marinheiro escritor*, 1866
 Tenho escrito um cento de cartas, feito
 have.IND.PRS.1SG write.PTCP DET.INDF.M.SG hundred of letters, do.PTCP
 outros tantos pedidos a pessoas que para lá vão; já
 other so.many pleas to persons that to there go.IND.PRS.3PL already
 mandei de propósito uma para tal fim, e **nada**
 send.PST.PFV.1SG of purpose DET.INDF.F.SG for such end and nothing
tenho podido conseguir
 have.IND.PRS.1SG be.able.PTCP achieve.INF
 ‘I have written some hundred letters, made even more pleas to people
 who are going there; I already sent one there on purpose to achieve this
 aim, and I haven’t been able to accomplish anything’

A very different situation can be observed in the verbal domain. Here, pluractionality is expressed by the use of adverbs or adverbials that express repetition or license a distributive interpretation, and the use of predicates that are compatible with the respective reading.

The adverbial expression *x vezes* (see example 17) is one of the most frequent ones to indicate iterativity and plays an important role in the strengthening of the reanalysis of the PPC as a form imbued with pluractional semantics (Becker 2020: 169, 181). However, it is a noteworthy fact that in the 19th century the PPC is still compatible with cardinal external quantification (see 18). It is not until the 20th century that the adverbial *x vezes* is only compatible with indefinite quantification, therefore, reflecting the feature of indeterminacy inherent to the semantics of the PPC.

- (17) Joaquim Manoel de Macedo, *O primo da Califórnia*, 1858
 também eles **têm-me recebido tantas vezes** em
 also they have.IND.PRS.3PL-me receive.PTCP so.many times in
 suas casas, que hoje por minha parte quero
 POSS.F.3PL houses that today from POSS.F.1SG part want.IND.PRS.1SG
 também recebê-los
 also receive.INF-them
 ‘Moreover, they have received me so many times in their houses that I now
 want to also receive them in mine’

- (18) Joaquim Manoel de Macedo, *Luxo e Vaidade*, 1860

já sei, tens tirado a sorte
 already know.IND.PRS.1SG have.IND.PRS.2SG draw.PTCP DET.DEF.F.SG luck
 grande cinco ou seis vezes.
 great five or six times
 'I (already) know, you have been lucky for five or six times.'

A pluractional reading in the verbal domain can also arrive from modification with a locative adverbial with distributive meaning. Thus, the locative adverbial in (19) licenses the inference that the announcement event must have taken place several times.

- (19) Júnior França, *As doutoras*, 1889

Não é aqui que mora uma doutora que
 not be.IND.PRS.3SG here that live.IND.PRS.3SG DET.INDF.F.SG doctor that
 tem anunciado nos jornais?
 have.IND.PRS.3SG announce.PTCP in.DET.DEF.M.PL newspapers
 'Doesn't here live a doctor who has placed ads in the newspapers?'

Finally, pluractionality in the verbal domain can arise on the basis of the type of predicate expressed by the verb. In particular, a pluractional interpretation is likely with durative predicates such as states (20) and atelic activities (21). In both cases, a durative reading arises. For instance, in (21) the question delimits a time interval that spans all of the life of the addressee, in which an event was repeated several times. Notably, such predicates co-occur more frequently with internal temporal modification (24 percent of all durative tokens) than other predicates (18 percent of all tokens from other predicate types). This difference reaches statistical significance ($\chi^2(1) = 3.95$, $p < 0.05^*$), which implies that examples such as (15) above are rather typical.

- (20) José de Alencar, *Mãe*, 1860

Tens tido notícias dele?
 have.IND.PRS.2SG have.PTCP news of.him
 'Have you received any news about him?'

- (21) Júnior França, *Caiu o ministério!*, 1883

Que empregos tem exercido?
 what work have.IND.PRS.2SGHON do.PTCP
 'What kind of work have you been doing?'

5 From nominal to verbal pluractionality

Having established a typology of the pluractional readings of the PPC and the contextual parameters associated with these readings in the nominal and verbal domains, we are now in a position to analyze the historical distribution of the PPC in terms of these contextual parameters. As argued in §2, the hypothesis that the semantic reanalysis of the PPC involved a transfer of the pluractional semantics from the nominal to the verbal domain makes the prediction that the use of the PPC was more frequent in nominal pluractionality contexts in earlier texts, whereas over time, its use came to be preferred in verbal pluractionality contexts.

As was already suggested in §2, sentences involving the PPC can be ambiguous between a resultative and a pluractional reading, and it is exactly these bridging contexts that are assumed to be at the heart of the semantic reanalysis of the PPC as a marker of pluractionality. Indeed, such ambiguous cases are a necessary part of any semantic change. As a result, a seemingly direct approach towards modeling the semantic change experienced by the PPC, namely annotating by semantic function as was done in Barbosa's (2014) study, is prone to subjectivity. In other words, it is possible that another researcher might come to different results regarding the annotation of the same data (cf. Rosemeyer 2016b for discussion).

Consequently, we adopted an indirect approach to the annotation of pluractionality in our data. Thus, we established aggregate variables for nominal and verbal pluractionality on the basis of a bottom-up categorization of each sentence in the data according to the contextual parameters identified as favoring these interpretations. Table 2 and 3 summarize the annotation processes for these two variables, *NOMINALPLURACTIONALITY* and *VERBALPLURACTIONALITY*.

Table 4 describes the distribution of the resulting numerical variables *NOMINALPLURACTIONALITY* and *VERBALPLURACTIONALITY*. It demonstrates that nominal pluractionality contexts are less frequent than verbal pluractionality contexts, in that 77.4 percent of all tokens of the PPC received a score of 0 on the *NOMINALPLURACTIONALITY* variable, whereas only 52.5 percent of all tokens of the PPC received a score of 0 on the *VERBALPLURACTIONALITY*.

Having established the operationalization of the variables *NOMINALPLURACTIONALITY* and *VERBALPLURACTIONALITY*, we proceeded to analyze the changes in the historical distribution of the PPC according to these two variables. Figure 3 visualizes this distribution. High values of *NOMINALPLURACTIONALITY* (upper plot) are frequent in 19th century plays, where around 25 percent of all PPC tokens reach a score of at least one on this variable. In contrast, after the 20th

Table 2: Summary of the annotation process for NOMINALPLURATIONALITY

Condition	Transformation for NOMINALPLURATIONALITY
Subject inflected for plural ($n = 91$)	+1
Theme refers to a distributive or plural referent ($n = 112$)	+1
Indirect object refers to a distributive or plural referent ($n = 13$)	+1

Table 3: Summary of the annotation process for VERBALPLURATIONALITY

Condition	Transformation for VERBALPLURATIONALITY
Presence of an aspectual or temporal adverbial that expresses internal or external quantification ($n = 198$)	+1
Presence of a locative adverbial that implies quantification ($n = 13$)	+1
Durative predicate type ($n = 349$)	+1

Table 4: Distribution of the variables NOMINALPLURATIONALITY and VERBALPLURATIONALITY

	0	1	2	3
NOMINALPLURATIONALITY	581	143	24	3
VERBALPLURATIONALITY	394	346	11	0

century, mean NOMINALPLURACTIONALITY decreases notably. This linear trend reached statistical significance.⁷

For VERBALPLURACTIONALITY, the overall trend is somewhat more complex. In particular, we find an increase between 1830 and 1929, followed by a sharp decrease between 1930 and 1949, which is in turn again followed by an increase. Again, this distribution reaches statistical significance.⁸

The historical distribution of the PPC in terms of the variables NOMINALPLURACTIONALITY and VERBALPLURACTIONALITY seems to confirm Amaral & Howe's (2012) hypothesis that the semantic change towards pluractionality involved a transfer from nominal to verbal pluractionality. In particular, we find that nominal pluractional contexts are especially frequent in earlier plays. Figure 3 suggests that as the frequency of verbal pluractional contexts increases, the frequency of nominal pluractional contexts decreases. However, there is one section of the data that does not seem to conform to this interpretation, namely the time period between 1930 and 1990. In this period, verbal pluractionality is less frequent than at the beginning of the 20th century. The nonlinearity of the semantic change of the PPC strongly suggests that, apart from purely semantic factors, sociolinguistic factors may have been at work. As we shall see in §6, one candidate parameter that might explain this distribution is register.

6 The role of register

Recent approaches to modeling historical change using diachronic corpus data have demonstrated that apparent gaps and inconsistencies in historical trends are frequently the result of changes in the textual corpus from which the analyzed data is taken. For instance, in his analysis of the historical development of the Brazilian Portuguese system of *wh*-interrogatives in data taken from the PorThea corpus, Rosemeyer (2019a) documents an increase in the usage frequency of bare *wh*-interrogatives such as *O que?* 'What?'. However, he argues that the use of

⁷Statistical significance was tested using an ordinal logistic regression model (Johnson & Albert 2004, Agresti 2010) predicting NOMINALPLURACTIONALITY from YEAR. The effect of YEAR reached statistical significance (log odds = -0.005, standard error = 0.002, $p < 0.05^*$). The modeling was realized in R (R Core Team 2019), using the ordinal package (Christensen 2019). We only tested for global significance of the trend, not significance of components.

⁸Statistical significance was tested using an ordinal logistic regression model predicting VERBALPLURACTIONALITY from YEAR. YEAR was modeled as a third-degree polynomial in order to capture the nonlinearity of the trend visualized in Figure 3. Two of the three components of the third-degree polynomial reached statistical significance (log odds_{poly1} = 4.733, standard error_{poly1} = 2.031, $p_{poly1} < 0.05^*$; log odds_{poly2} = 1.141, standard error_{poly2} = 1.026, $p_{poly2} > 0.05$; log odds_{poly3} = 6.981, standard error_{poly3} = 3.432, $p_{poly3} < 0.001^{***}$). Cf. also Table 5 in §6.

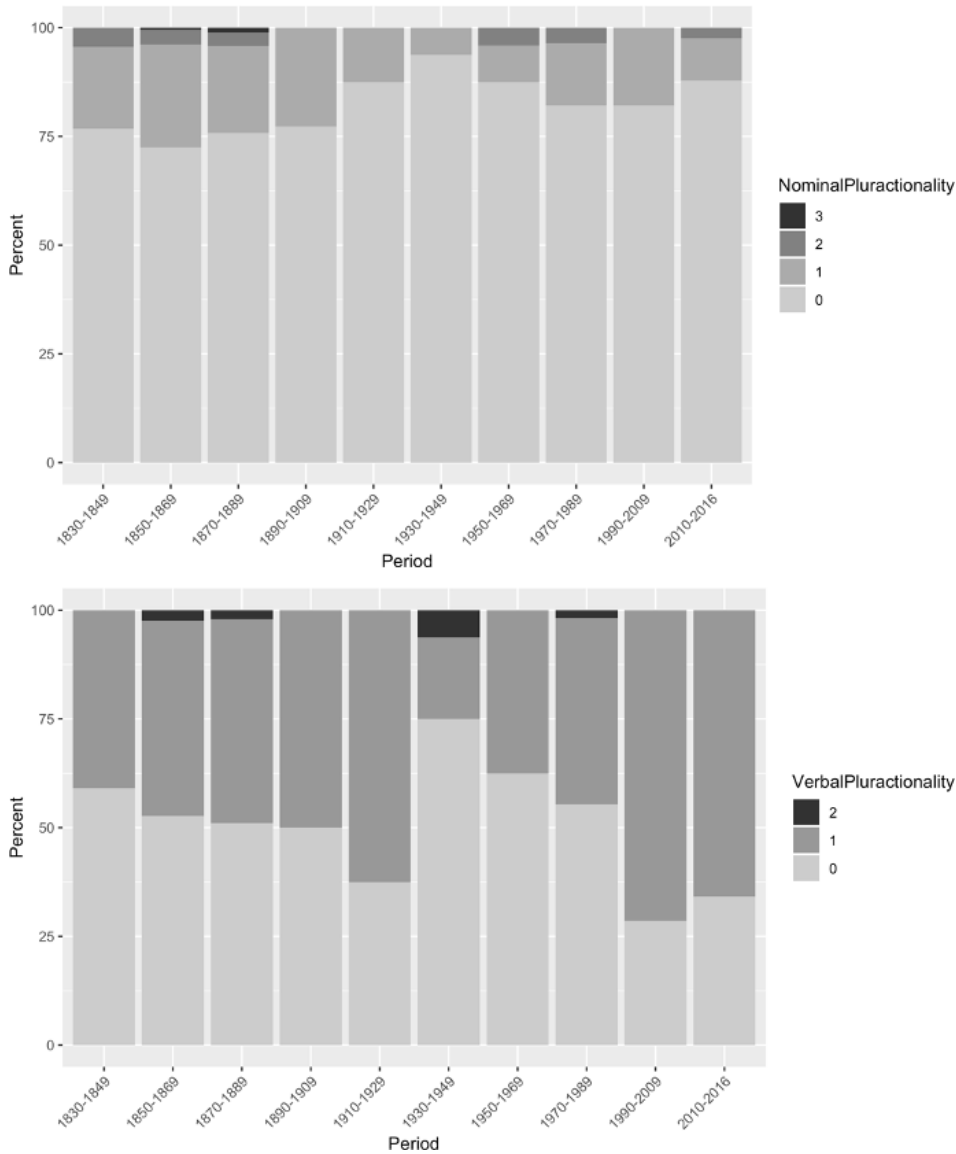


Figure 3: Scores on the NOMINALPLURATIONALITY and VERBALPLURATIONALITY variables, by 20-year periods

such bare interrogatives is unlikely to undergo any major semantic or syntactic changes, as bare interrogatives are always strongly dependent on the previous context (e.g., the question *Onde?* ‘where’ needs an antecedent in order for the hearer to be able to reconstruct a complete proposition such as *Onde a gente viu a Maria?* ‘Where did we see Maria?’). Bare interrogatives are shown to be more typical for spoken language and texts that can be characterized as conceptually oral, i.e. approximating spoken language to a greater degree. Consequently, Rosemeyer tests the hypothesis that the increase in the usage frequency of bare interrogatives is due to a genre change in the corpus of theater plays; over time, theater plays have come to approximate orality to a greater degree, which is why bare interrogatives are used more frequently. Such a genre change can also be described as a change from more formal to less formal register.

Rosemeyer (2019a: 175) establishes a bottom-up measurement of the degree to which the theater plays approximate orality in the PorThea corpus. On the basis of the operationalization of the “involvement” dimension established in Biber & Finegan (2004), the orality variable was defined as the joint log-transformed normalized usage frequency of a number of linguistic variables, namely (a) “private” verbs in present tense singular (e.g. *achar* ‘to mean’ and *pensar* ‘to think’ etc.), (b) present progressives, (c) demonstrative neuter pronouns (*isso* and *isto* ‘this’), (d) time and place adverbs (*aqui* ‘here’ and *agora* ‘now’) and (e) discourse markers such as *bom* ‘well’ or *pois* ‘so’). We assume with Biber & Finegan that the use of these linguistic variables is typical for conceptually oral texts. The higher the score of a text for the resulting variable LOGORALITY, the more a text is expected to approximate orality.

By controlling for LOGORALITY while describing the changes in the usage frequency of bare interrogatives in the PorThea corpus, Rosemeyer (2019a) is able to demonstrate that the frequency increase disappears when register is taken into account. In other words, the increase in the usage frequency of bare interrogatives is an “apparent change” that depends entirely on the composition of the textual corpus.

In order to investigate the possibility that the semantic change of the PPC towards verbal pluractional readings was an apparent change, we applied Rosemeyer’s (2019a) variable LOGORALITY to our data. Consequently, we were able to investigate whether the semantic change of the PPC towards verbal pluractional readings represents an actual change (in the sense that it is not due to changes in register over time) or apparent change (in the sense that the change is explained by the historical development of the genre of theater texts).

In addition, applying the variable LOGORALITY to our data allowed us to gauge to which degree the semantic change towards verbal pluractionality readings is

moderated by register. This is important because it gives an understanding of whether the change originated in texts with a high degree of conceptual orality or in texts with a low degree of conceptual orality. In the first case, the change can be classified as a “change from below”, i.e. a change that seems to have originated in spoken language first and then spread to more formal registers. In the second case, the change would represent a “change from above”, i.e. a change that originated in more formal registers and then spread to lower registers.

Let us first turn to the question of whether the semantic change towards pluractionality constitutes actual or apparent change. As was mentioned in the description of the case study of bare interrogatives in Rosemeyer (2019a), in order for a change to be “explained away” by register differences, there needs to be a correlation between the variable whose distribution is being analyzed and the degree of conceptual orality. Only a weak marginally significant correlation was found for the variables `VERBALPLURACTIONALITY` and `LOGORALITY`.⁹ As a result, it seems unlikely that the change towards the expression of verbal pluractionality found in the data is an artifact of the composition of the corpus.

In contrast, our analysis did find evidence for the assumption that the semantic change towards pluractionality was moderated by the degree to which the theater plays approximate spoken language. We expanded the ordinal logistic regression predicting `VERBALPLURACTIONALITY` from `YEAR`, whose results were presented in §5 (see footnote 7), by including an interaction effect between `YEAR` and `ORALITYHIGH`. The variable `ORALITYHIGH` assumed a value of “False” for low orality texts (where the value of `LOGORALITY` was below the mean value of `LOGORALITY`) and “True” for high orality texts (where the value of `LOGORALITY` was above the mean value of `LOGORALITY`). Table 5 summarizes the results from this model, including the original baseline model for comparison.

As evident in Table 5, the interaction between `YEAR` and `ORALITYHIGH` reached statistical significance (cf. the last three lines in the table), which means that the degree of conceptual orality of the theater plays indeed significantly moderated the semantic change towards verbal pluractionality.

A pair-wise model comparison using ANOVA (see Gries 2013: 285–293 for application) found the extended model to explain significantly more variation than the baseline model, which also justifies inclusion of `ORALITYHIGH` as a main and interaction effect.

Figure 4 visualizes the historical distribution of verbal pluractionality when distinguishing between low and high orality texts, resulting from the interac-

⁹Correlation testing was done using Kendall’s τ because both variables were not normally distributed (Gries 2009: 213). No significant correlation between `LOGORALITY` and `VERBALPLURACTIONALITY` was found (Kendall’s $\tau = -0.039$, $z = 1.317$, $p_{\text{one-tailed}} < 0.1$).

Table 5: Summary of results from the two ordinal logistic regression models measuring the correlation between VERBALPLURACTIONALITY, YEAR, and ORALITYHIGH. Abbreviations: LO = Log odds, SE = standard error. YEAR was modeled as a third-degree polynomial using the function `poly()` in order to account for the fact that the increase in verbal pluractionality is a non-linear trend.

Variable	Baseline model				Extended model			
	LO	SE	<i>z</i>	<i>p</i>	LO	SE	<i>z</i>	<i>p</i>
Main effects								
poly(YEAR, 3)1	4.7	2.0	2.3	<0.054	0.1	3.4	0.0	>0.05
poly(YEAR, 3)2	1.1	2.0	0.6	<0.05	-2.7	2.9	-1.0	>0.05
poly(YEAR, 3)3	7.0	2.0	3.4	<0.001	10.4	3.2	3.2	<0.01
ORALITYHIGH					-0.3	0.2	-1.7	<0.1
Interaction effects								
poly(YEAR, 3)1 ORALITYHIGH					6.9	4.4	1.6	>0.05
poly(YEAR, 3)2 ORALITYHIGH					9.7	4.4	2.2	<0.05
poly(YEAR, 3)3 ORALITYHIGH					-7.6	4.3	-1.8	<0.1

tion between YEAR and ORALITYHIGH in the extended ordinal logistic regression model described in Table 5.¹⁰ It demonstrates significant differences in the semantic change of the PPC towards verbal pluractionality in texts scoring low or high on the dimension of conceptual orality. In particular, the curious “two-wave” distribution of the development towards verbal pluractionality documented in §5 (see Figure 3) can be explained in terms of register differences. Thus, the first increase in verbal pluractionality, between about 1840 and 1880, is restricted to low-orality texts; in high-orality texts we do not find an increase in verbal pluractionality contexts until 1950. However, between 1880 and 1975 the trend towards verbal pluractionality in low-orality texts is actually inverted, such that the PPC is used less in verbal pluractionality contexts, evening out the differences between low- and high-orality texts. After 1950 (1975 for low-orality texts), we find a relatively uniform increase in verbal pluractionality in both low- and high-orality texts. In present-day theater plays, the degree of conceptual orality does not seem to affect the frequency of verbal pluractionality contexts.

¹⁰The effect plots were produced using the effects package (Fox & Hong 2009).

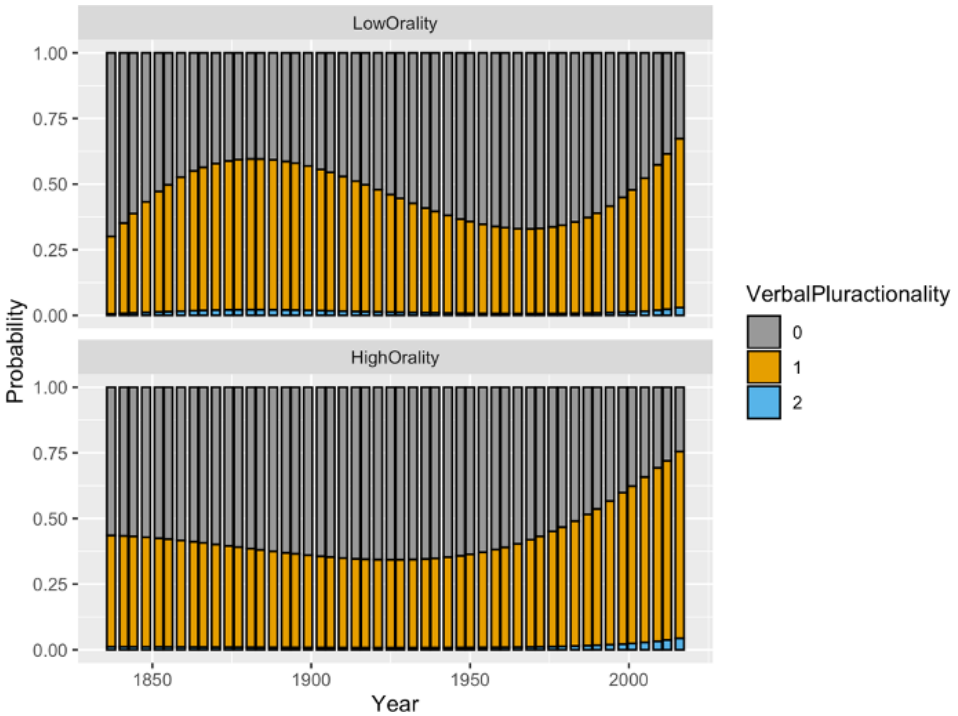


Figure 4: Effect plot for the interaction between YEAR (as 10-year periods) and HIGHORALITY in the ordinal logistic regression model predicting VERBALPLURACTIONALITY

7 Discussion and conclusions

In this paper, we have modeled the semantic change of the Brazilian Portuguese PPC from resultative to pluractional readings. We tested the hypothesis that nominal pluractionality contexts served as bridging contexts for this change, leading to a transfer of the pluractionality reading from nominal to verbal properties. Using quantitative data from a corpus of BP theater plays, we have been able to verify this hypothesis. Our data clearly shows a preference for the PPC to appear in contexts associated with nominal pluractionality in earlier stages of the change. Over time, its use became more likely in contexts associated with verbal pluractionality. The results from this paper thus confirm assumptions from previous studies based on qualitative analyses.

Closer inspection of the trend in terms of usage frequencies, type-token ratios and the degree of conceptual orality of a text has revealed that it proceeded in two qualitatively different phases. Thus, the PPC came to be associated more strongly with verbal pluractionality contexts in a first phase between 1840 and

1880. However, this change was restricted to low-orality texts, which correspond to a more formal register. Note also that the semantic change was correlated to a decreasing usage frequency of the PPC. This is surprising given that grammaticalization processes are usually expected to involve an increase in usage frequency (grammaticalization involves an extension of the use of the construction to new usage contexts, which leads to an increase in overall usage frequency). Indeed, inspection of the development of the type-token ratios of the PPC per year demonstrated that the mid-19th century marks the beginning of the extension of the use of the PPC to new verb types and consequently, an increase in productivity.

How can we explain the fact that the PPC decreased in frequency at the same time that its productivity increased due to the semantic change towards pluractionality? One possible explanation, which was already alluded to at the end of §2, is contact. Consider the periodization of Brazilian Portuguese established in Galves (2007). Galves provides a summary of historical studies on several aspects of the grammar of European and Brazilian Portuguese. On the basis of this summary, she claims that BP grammar changed significantly with respect to EP in the first two centuries of colonization, which is why some studies find evidence for the emergence of a new grammar already in 18th century texts.

However, there is also evidence for a revival of EP influence on BP grammar in the 19th century. Galves quotes the studies by Carneiro (2005) and Pagotto (1992), which analyze the historical distribution of pronoun position (enclitic vs. proclitic). While BP had changed from preferred enclitic to proclitic position during the 19th century, these authors document a return to preferred enclitic position. The three authors agree in attributing this change to the strong cultural influence of Portugal at the time of the so-called *Império do Brasil* (1822–1889) and the first decades of the *República Velha* (from 1889 onwards). Official documents (such as the Constitution of 1891 and other administrative documents, but also literary prose, aspired to imitate and even to exceed the *norma culta* of the European Portuguese (see Pagotto 1998: 51–53). The huge impact of the European Portuguese norm is evident especially in the text of the First Republican Constitution of 1891, which Pagotto (1992, 1998) compared with the Constitution of the Empire of 1824 (*Constituição do Império*). He concludes that the 1824 Constitution, in contrast to the 1891 Constitution, favors proclitic pronoun position and is, therefore, still closer to the classical Portuguese norm. In contrast, the 1891 Constitution clearly prefers enclitic pronoun position and thus consistently follows the European norm of the time, which tends towards a generalization of the enclitic pronominal position even in contexts where the proclisis was still common in classical language (Pagotto 1998: 51–53).

Our data documenting the semantic evolution of the PPC seems to evince a similar tendency of increasing EP influence. As we have seen, several findings from this paper support the hypothesis that the semantic change of the PPC towards expression of pluractionality was fostered by the influence of EP grammar during the 19th century. First, this hypothesis requires as a premise that the PPC displays a stronger tendency to express pluractionality in EP than in BP. The periodization analysis (see §3) may be interpreted as evidence for this premise. In particular, the usage frequency of the PPC is significantly higher in 20th century EP texts than in BP texts, which might suggest that the semantic change has been implemented to a greater degree in EP than in BP. Second, such language contact is unlikely to have affected low register and informal language. Rather, we would expect this change to affect more formal language, which is what we find in this study. In particular, we only document a semantic change of the PPC towards pluractionality in more formal theater plays, i.e. plays that do approximate the language spoken in Brazil at that time. Note also that after the end of the *Império do Brasil*, our data actually suggests a “de-pluractionalization” of the PPC in formal theater plays; in the first half of the 20th century, authors seem to have reverted to the Brazilian Portuguese norm of using the PPC. Third, the assumption of a contact-induced change towards pluractionality is compatible with the finding of an overall decrease in the usage frequency of the PPC. Figure 5 uses the orality measure established in §6 to model more finely the development of usage frequencies in our 19th century data. It demonstrates that the decrease in the usage frequency of the PPC in our 19th century data is actually restricted to high-orality texts; in low-orality texts, the usage frequency of the PPC remains roughly similar until about 1880. It is only after 1880 that the use frequency of the PPC starts to decrease also in low-orality texts. Note that the year 1880 was also identified as the turning point with respect to the semantic change towards pluractionality in low-orality texts (see the discussion of Figure 4 in §5), in that after the end of the 19th century, verbal pluractionality readings actually became less frequent in low-orality texts. These findings seem to fit an explanation in terms of contact between EP and BP quite well.

Let us now turn to the second phase of the semantic change, i.e. the increase in verbal pluractionality readings after 1950. This increase in verbal pluractionality was not found to be moderated by the degree of conceptual orality of the texts; in other words, the likelihood for the PPC to be used in contexts associated with verbal pluractionality readings increases both in low- and high-orality texts. At the same time, however, we find a decrease in the usage frequency of the PPC, as well as its type-token ratio, which strongly suggests a general decrease in productivity of the construction. Consequently, it appears that after about 1950, the

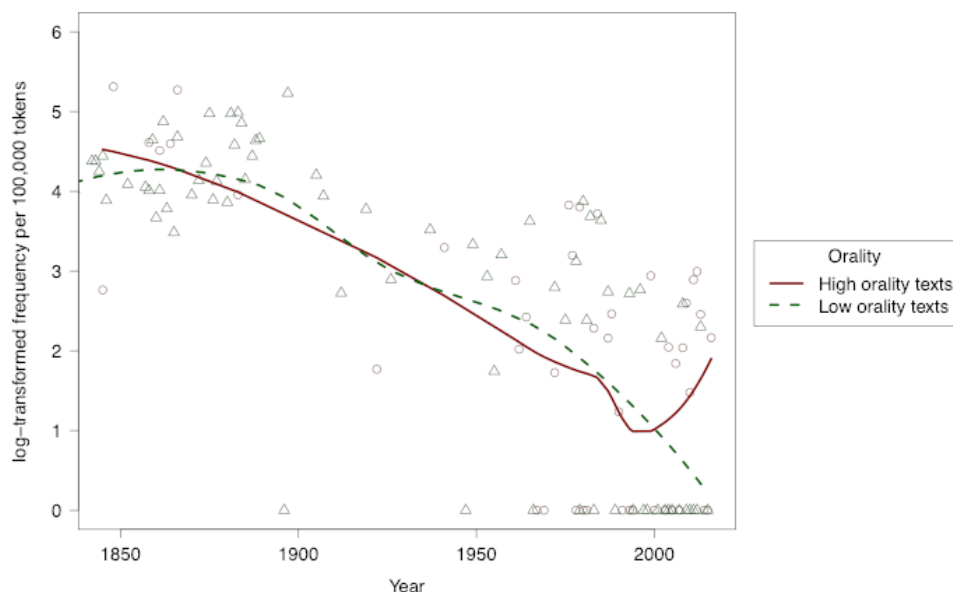


Figure 5: Historical development of the log-transformed usage frequency per 100,000 tokens of the PPC in the Brazilian section of the PorThea corpus, by orality. Points represent mean usage frequencies per year, whereas the line represents results of local polynomial regression analyses summarizing the trend.

PPC has experienced a specialization process by which its use has gradually been restricted to contexts that are strongly associated to verbal pluractional readings. The restriction of constructions to such functional niches is a hallmark of conservation processes, for instance in situations of language change in which one construction is being replaced by another, competing construction (see Rosemeyer 2016a). In this case, it stands to reason that the gradual specialization of the PPC is due to the competition with the simple past (henceforth PPS). According to the variationist analysis by Barbosa (2014), already cited in §2, the PPS is gradually ousting the PPC in BP. Thus, Barbosa's (2014) data suggest an increase of the frequency of the PPS relative to the PPC of 85.6 percent in the 18th century to 89.2 percent in the 19th century and an almost-categorical 94.8 percent in the 20th century. Another hallmark of historical replacement processes is that the usage of the competing construction can even end up expanding to those functional niches in which the replaced construction seems to still thrive. There is some evidence for the assumption that in BP the PPS is starting to be used in verbal pluractionality contexts, a change that might lead to the complete elimination of the PPC from BP grammar. The examples in (22–23) are taken from BP texts

from the News on the Web (=NOW) section of the CdP, which includes 1.1 billion words from internet news texts dated between 2012 and 2019. In these examples, the PPS is used in contexts that clearly indicate verbal pluractionality and where, according to BP grammars, use of the PPC would be expected.

(22) BP examples of the syntagm *fiz até agora* from the CdP, section NOW

- a. Eu avalio com muita felicidade e gratidão tudo
I assess.IND.PRS.1SG with much happiness and gratitude everything
o que **fiz** até **agora**
DET.DEF.M.SG that do.IND.PST.PFV.1SG until now
‘I see everything I have achieved until now with much happiness and gratitude’
- b. Venho apresentar minha defesa e dizer a
come.IND.PRS.1SG present my defense and say.INF DET.DEF.F.SG
verdade, como sempre **fiz** até **agora**
truth, like always do.IND.PST.PFV.1SG until now
‘I will now present my defense and tell the truth, as I have always done until now’
- c. o terceiro álbum está a caminho e
DET.DEF.M.SG third album be.PRS.3SG to way and
foi o melhor que já **fiz**
be.PST.PFV.3SG DET.DEF.M.SG best that already do.IND.PST.PFV.1SG
até agora
until now
‘The third album is on its way and it was [sic] the best that I have done until now’

(23) BP examples of the syntagm *vivi até agora* from the CdP, section NOW

- a. Deram-me um ano de vida, exatamente o
give.IND.PST.PFV.3PL me one year of life exactly DET.DEF.M.SG
que **vivi** até **agora**
that live.PST.PFV.1SG until now
‘They gave me one [remaining] year to live, which is exactly how long I have lived until now’

- b. “Talvez eu tenha mais vida para viver do
 maybe I have.SBJ.PRS.1SG more life to live.INF of.DET.DEF.M.SG
 que eu já **vivi** **até agora**”, brinca
 that I already live.PST.PFV.1SG until now joke.ind.prs.3sg
 “‘Maybe I will have more life to live than I have lived until now’, s/he
 jokes’

Table 6 summarizes the distribution of the distribution of the PPS and the PPC in the contexts of the adverbial *até agora* (‘until now’) in the NOW section of the CdP.

Table 6: Summary of the distribution of the PPS and the PPC in the contexts of the adverbial *até agora* (‘until now’) in the NOW section of the CdP. Percentages refer to the relative frequencies of the PPS and PPC within each dialect.

Query/Type	BP		EP		Total
	<i>n</i>	%	<i>n</i>	%	
_vis% até agora (= PPS)	2974	90.0%	2271	76.9%	5245
TER_vps* até agora (= PPC)	297	9.1%	684	23.1%	981
Total	3271		2955		
$\chi^2(1) = 230.39, p < 0.001^{***}$					

For the BP section of the data, Table 6 demonstrates that in these contexts, which strongly suggest pluractional readings, the PPS is used in more than 90 percent of the cases, with the PPC relegated to a clear minority variant. While the overall pattern of the distribution is similar in EP, the asymmetry between the PPS and the PPC is less marked given that the PPC is used in about 23 percent of the cases. This finding, which reaches significance according to a χ^2 test, suggests that the PPC is more strongly established in verbal pluractionality contexts in EP than in BP and seems to resist replacement with the PPS to a greater degree than in EP.

References

- Agresti, Alan. 2010. *Analysis of ordinal categorical data*. 2nd edn. Hoboken: Wiley.
10. DOI: 10.1002/9780470594001.
- Amaral, Patrícia & Chad Howe. 2012. Nominal and verbal plurality in the diachrony of the Portuguese present perfect. In Patricia Cabredo Hofherr & Brenda Laca (eds.), *Verbal plurality and distributivity*, 25–54. Berlin: De Gruyter.
- Barbosa, Juliana Bertucci. 2008. *Tenho feito / Fiz a tese: Uma proposta de caracterização do Pretérito Perfeito no Português*. Universidade Estadual Paulista, Faculdade de Ciências e Letras, Campus de Araraquara. (Doctoral dissertation). <http://hdl.handle.net/11449/103584>.
- Barbosa, Juliana Bertucci. 2012. O pretérito perfeito no português do Brasil e no português de Portugal contemporâneo. *Todas as Letras* 14(1). 176–194.
- Barbosa, Juliana Bertucci. 2014. Estudo diacrônico do pretérito perfeito no português do Brasil. In Daniel Soares da Costa (ed.), *Pesquisas linguísticas pautadas em corpora*, 85–110. São Paulo: Editora UNESP.
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic* (Constructional Approaches to Language 8). Amsterdam, Philadelphia: Benjamins.
- Becker, Martin. 2017. O pretérito perfeito composto em diacronia: Uma evolução perfeita? *Estudos de Lingüística Galega* 8. 25–43.
- Becker, Martin. 2020. O Pretérito Composto (PPC) na história do português do Brasil. In Rodolfo Ilari & Renato Basso (eds.), *História Semântica do português Brasileiro*, 146–203. São Paulo: Editora Contexto.
- Bertinetto, Pier Marco & Eugenio Cividini. 2015. The semantics of degree verbs and the telicity issue. *Borealis* 4(1). 57–77.
- Bertinetto, Pier Marco & Alessandro Lenci. 2012. Habituality, pluractionality, and imperfectivity. In Robert I. Binnik (ed.), *The Oxford handbook of tense and aspect*, 852–880. Oxford: Oxford University Press.
- Biber, Douglas & Edward Finegan. 2004. Historical drift in three English genres. In Geoffrey Sampson & Diana McCarthy (eds.), *Corpus linguistics: Readings in a widening discipline*, 67–77. London: Continuum.
- Boléo, Manuel de Paiva. 1936. O perfeito e o pretérito em português em confronto com as outras línguas românicas. *Cursos e Conferências, Vol. VI, Suplemento ao vol. XIII do Boletim da Biblioteca da Universidade*. 124–266.
- Bybee, Joan L & Rena Torres Cacoullos. 2009. The role of prefabs in grammaticalization: How the particular and the general interact in language change. In Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali & Kathleen M. Wheatley

- (eds.), *Formulaic language*, vol. 1: Distribution and historical change, 187–217. Amsterdam: Benjamins. DOI: 10.1075/tsl.82.09the.
- Cabredo Hofherr, Patricia, Brenda Laca & Sandra De Carvalho. 2010. When perfect means plural: The present perfect in Northeastern Brazilian Portuguese. In Patricia Cabredo Hofherr & Brenda Laca (eds.), *Layers of aspect*, 67–100. Stanford, CA: CSLI Publications.
- Carneiro, Zenaide Novaes. 2005. *Cartas da Bahia (1809-1907): Um estudo filológico-linguístico*. Campinas: Unicamp. (Doctoral dissertation).
- Christensen, Rune Haubo Bojesen. 2019. *ordinal: Regression models for ordinal data*. <https://CRAN.R-project.org/package=ordinal> (6 March, 2020).
- Collins, Jeremy. 2019. Some language universals are historical accidents. In Karsten Schmidtke-Bode, Natalya Levshina, Susanne Maria Michaelis & Ilja A. Seržant (eds.), *Explanation in typology: Diachronic sources, functional motivations and the nature of the evidence* (Conceptual Foundations of Language Science 3), 47–61. Berlin: Language Science Press. DOI: 10.5281/zenodo.2583807.
- Cornillie, Bert & Bridget Drinka. 2019. Latin influence on the syntax of the languages of Europe: Foundations and new perspectives. *Belgian Journal of Linguistics* 33. DOI: 10.1075/bjl.00020.cor.
- Cristofaro, Sonia. 2019. Taking diachronic evidence seriously: Result-oriented vs. source-oriented explanations of typological universals. In Karsten Schmidtke-Bode, Natalya Levshina, Susanne Maria Michaelis & Ilja A Seržant (eds.), *Explanation in typology: diachronic sources, functional motivations and the nature of the evidence* (Conceptual Foundations of Language Science 3), 25–46. Berlin: Language Science Press. DOI: 10.5281/zenodo.2583805.
- Davies, Mark. 2016–. *O corpus do português*. <https://www.corpusdoportugues.org> (2 April, 2018).
- Fox, John & Jangman Hong. 2009. Effect displays in R for multinomial and proportional-odds logit models: Extensions to the effects package. *Journal of Statistical Software* 32(1). 1–24.
- Galves, Charlotte. 2007. A língua das caravelas: Periodização do português europeu e origem do português brasileiro. In Ataliba De Castilho, Maria Aparecida Torres Moraes, Ruth Lopes & Sonia Cyrino (eds.), *Descrição, história e aquisição do português brasileiro*, 513–528. Campinas: Pontes.
- Gärtner, Eberhard. 1998. *Grammatik der portugiesischen Sprache*. Berlin: De Gruyter.
- Gries, Stefan Thomas. 2009. *Quantitative corpus linguistics with R: A practical introduction*. New York, London: Routledge.
- Gries, Stefan Thomas. 2013. *Statistics for linguistics with R: A practical introduction*. 2nd edn. Berlin: De Gruyter.

- Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer & Gabriele Diewald (eds.), *New reflections on grammaticalization*, 83–101. Amsterdam: Benjamins. DOI: 10.1075/tsl.49.08hei.
- Himmelmann, Nikolaus P. 2004. Lexicalization and grammaticization: Opposite or orthogonal? In Walter Bisang, Nikolaus P. Himmelmann & Björn Wiemer (eds.), *What makes grammaticalization? A look from its fringes and its components*, 21–42. Berlin: De Gruyter.
- Hundertmark-Santos Martins, Maria Teresa. 1982. *Portugiesische Grammatik*. Tübingen: Niemeyer.
- Ilari, Rodolfo. 1999. *Linguística românica*. 3rd edn. São Paulo: Ática.
- Ilari, Rodolfo. 2001a. *A expressão do tempo em português*. 2nd edn. São Paulo: Editora Contexto.
- Ilari, Rodolfo. 2001b. Notas sobre o passado composto em português. *Revista Letras (Curitiba)* 55. 129–152.
- Johnson, Valen E. & James H. Albert. 2004. Ordinal regression models. In David Kaplan (ed.), *The SAGE handbook of quantitative methodology for the social sciences*, 151–174. Thousand Oaks: SAGE.
- Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz: Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgebrauch. *Romanistisches Jahrbuch* 36. 15–43.
- Kytö, Merja. 2011. Corpora and historical linguistics. *Revista Brasileira de Linguística Aplicada* 11(2). 417–457.
- Laca, Brenda. 2006. Indefinites, quantifiers and pluractionals: What scope effects tell us about event pluralities. In Liliane Tasmowski & Svetlana Vogeleer (eds.), *Non-definiteness and plurality*, 191–217. Amsterdam: John Benjamins.
- Laserson, Peter. 1995. *Plurality, conjunction and events*. Kluwer: Dordrecht.
- Leonetti, Manuel. 2017. Basic constituent orders. In Andreas Dufter & Elisabeth Stark (eds.), *Manual of Romance morphosyntax and syntax*, 887–932. Berlin: De Gruyter.
- McEnery, Anthony Mark & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- Michaelis, Laura A. 2004. Type shifting in construction grammar: An integrated approach to aspectual coercion. *Cognitive Linguistics* 15. 1–67.
- Molsing, Karina Veronica. 2010. *The present perfect: An exercise in the study of events, plurality and aspect*. Pontifícia Universidade Católica do Rio Grande do Sul. (Doctoral dissertation).
- Novaes, Celso Vieira & Juana Barros Nespole. 2014. O traço aspectual de perfect e as suas realizações. *Revista FSA* 11(1). 255–279.

- Olbertz, Hella. 2018. The perfect in (Brazilian) Portuguese: A Functional Discourse Grammar view. *Open Linguistics* 4. 478–508.
- Oliveira, Fátima & António Leal. 2012. Sobre a iteração do Pretérito Perfeito Composto em português Europeu. *Revista de Estudos Linguísticos da Univerdade do Porto* 7. 65–88.
- Pagotto, Emílio. 1992. *A posição dos clíticos em português, um estudo diacrônico*. (MA thesis). Campinas.
- Pagotto, Emílio. 1998. Norma e condescendência: Ciência e pureza. *Línguas e Instrumentos Linguísticos* 2. 49–68.
- Perini, Mário. 2002. *Modern Portuguese: A reference grammar*. New Haven: Yale University Press.
- Poplack, Shana, Lauren Zentz & Nathalie Dion. 2011. What counts as (contact-induced) change. *Bilingualism: Language and Cognition* 15(02). 247–254. DOI: 10.1017/S1366728911000502.
- Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, Mass.: MIT Press.
- R Core Team. 2019. *R: A language and environment for statistical computing*. <http://www.R-project.org> (15 November, 2017).
- Rosemeyer, Malte. 2016a. Modeling frequency effects in language change. In Heike Behrens & Stefan Pfänder (eds.), *Experience counts: Frequency effects in language*, 175–207. Berlin: De Gruyter.
- Rosemeyer, Malte. 2016b. The development of iterative verbal periphrases in Romance. *Linguistics* 54(2). 235–272.
- Rosemeyer, Malte. 2019a. Actual and apparent change in Brazilian Portuguese wh-interrogatives. *Language Variation and Change* 31(2). 165–191.
- Rosemeyer, Malte. 2019b. PorThea: A historical corpus of Portuguese theater plays.
- Santos, Diana. 2008. Perfect mismatches: ‘Result’ in English and Portuguese. In Gunilla M. Anderman & Margaret Rogers (eds.), *Incorporating corpora: The linguist and the translator*, 217–242. Clevedon: Multilingual Matters.
- Schmitt, Christina. 2010. Cross-linguistic variation and the present perfect: The case of Portuguese. *Natural Language and Linguistic Theory* 19. 403–453.
- Szmrecsanyi, Benedikt. 2016. About text frequencies in historical linguistics: Disentangling environmental and grammatical change. *Corpus Linguistics and Linguistic Theory* 12(1). 153–171.
- Thomason, Sarah G. 2010. Contact explanations in linguistics. In Raymond Hickey (ed.), *The handbook of language contact*, 31–47. London: Wiley-Blackwell.

- van der Auwera, Johan & Volker Gast. 2012. What is “contact induced grammaticalization”? Examples from Mayan and Mixe-Zoquean languages. In Björn Wiemer, Bernhard Wälchli & Björn Hansen (eds.), *Grammatical replication and borrowability in language contact*, 381–426. Berlin: De Gruyter.
- von Stechow, Arnim. 1999. Eine erweiterte Extended Now-Theorie für Perfekt und Futur. *Zeitschrift für Literaturwissenschaft und Linguistik* 113. 86–118.

Name index

- Agresti, Alan, 243
Albert, James H., 243
Alcover, Joan, 144–147, 150, 154
Almeida, Maria Clotilde, 119
Aloni, Maria, 2, 6, 48, 63, 65, 66, 142, 163
Alonso-Ovalle, Luis, 2, 48, 67, 69, 142
Amaral, Patrícia, 9, 221–230, 243
Antolí Martínez, Jordi M., 148
Asis, Jorge, 81
Avram, Mioara, 17, 25, 35, 38
- Badia i Margarit, Antonio Maria, 147
Baily, Samuel L., 46, 47
Balsadella, Davide, 209
Barbosa, Juliana Bertucci, 225–227, 231, 233, 241, 251
Barðdal, Jóhanna, 233
Batlle, Mar, 146, 147
Becker, Martin, 143, 145, 225–230, 239
Beghelli, Filippo, 178
Bertinetto, Pier Marco, 226, 237
Bertocchi, Alessandra, 184
Biber, Douglas, 245
Bickerton, Derek, 103, 105, 106
Biyoko Mabua, Jean-Claude Simon, 117
Blasco Ferrer, Eduardo, 178–181, 185, 186, 188, 205
Bloch, Oscar, 145
Boléo, Manuel de Paiva, 225
- Bontinck, François, 118
Boot, Teodoro, 79
Borkowska, Paulina, 72
Bosque, Ignacio, 93, 106, 124
Bostoen, Koen, 89, 116, 118
Brero, Camillo, 46
Brucart, Joseph M., 145, 153, 154
Bybee, Joan L, 233
- Cabredo Hofherr, Patricia, 225, 226
Cacho Blecua, Juan Manuel, 195
Cameron-Faulkner, Thea, 124
Campos, Ednalvo Apóstolo, 107
Camus Bergareche, Bruno, 6, 8
Cancellier, Antonella, 47
Caponigro, Ivano, 165
Cardinaletti, Anna, 3, 14, 38, 41
Cardoso, Hugo C., 121
Carlier, Anne, 3, 6
Carlson, Greg N., 23, 45
Carneiro, Zenaide Novaes, 249
Carter, Hazel, 117
Chaudenson, Robert, 89, 108
Chicuna, Alexandre Mavungo, 96
Chierchia, Gennaro, 2, 48, 49, 67, 68
Chiriacescu, Sofiana, 20, 21, 40, 41
Choi, Jinyoung, 163
Chomsky, Noam, 211
Christensen, Rune Haubo Bojesen, 243
Cinque, Guglielmo, 169
Cirillo, Robert J., 209–211, 214

Name index

- Civardi, Eugenio, 237
Cocco, Maurizio, 56
Collins, Jeremy, 231
Colomina i Castanyer, Jordi, 154
Company-Company, Concepción, 6,
8, 141, 143, 145, 146, 148, 150,
154, 157, 165
Conde, Oscar, 46, 48
Cornillie, Bert, 231
Cornips, Leonie, 14
Coromines, Joan, 144, 146, 147, 153,
157
Cortellazzo, Manlio, 56
Corver, Norbert, 211, 212
Cristofaro, Sonia, 231

Davatz, Jan, 16, 23, 24, 26, 36, 38
Davies, Mark, 222, 228, 231
de Borja Moll, Francesc, 144–147, 150,
154
de Carvalho do Vale, Rosana
Siqueira, 107
de Friedemann, Nina S., 92, 94, 99
De Mauro, Tullio, 46
de Miguel, Elena, 4
de Schryver, Gilles-Maurice, 89, 116,
118
de Swart, Peter, 40
DeCamp, David, 103
Degano, Marco, 6
Demonte, Violeta, 4
Dereau, Léon, 117
Devoto, Fernando, 46
Di Tullio, Ángela Lucía, 46, 47, 52
Diewald, Gabriele, 75
Dobrovie-Sorin, Carmen, 4, 16, 18, 40
Doetjes, Jenny, 209
Dom, Sebastian, 116
Drinka, Bridget, 231

Engels, Kathrin, 46–48
Ennis, Juan Antonio, 47
Enrique-Arias, Andrés, 76, 149
Etxeberria, Urtzi, 66, 152

Fălăuș, Anamaria, 2, 67
Farina, Luigi, 182
Farkas, Donka, 18
Finegan, Edward, 245
Fodor, Janet D., 2
Fox, John, 247
Francia, Marika, 49, 160, 163, 167, 169
Franco, Irene, 189–191

Gabriel, Christoph, 95
Galves, Charlotte, 249
Gamillscheg, Ernst, 145
Gärtner, Eberhard, 225
Gast, Volker, 231
Gerards, David, 38
Geurts, Bart, 72
Giannakidou, Anastasia, 65, 66, 152,
187
Gianollo, Chiara, 6, 143, 187
Girón Alconchel, José Luis, 145
Giusti, Giuliana, 3, 14, 38, 41
Godefroy, Frederic, 145
Gonçalves, Perpétua, 128
Grandgent, Charles Hall, 194
Greimas, Algirdas Julien, 145
Gries, Stefan Thomas, 233, 246
Güldemann, Tom, 111, 114
Guthrie, Malcolm, 89
Gutiérrez Maté, Miguel, 90, 92, 100,
102, 105, 108, 109, 120, 129,
132, 133
Gutiérrez-Rexach, Javier, 93
Gutzmann, Daniel, 79

Haensch, Günther, 46, 50

- Hardie, Andrew, 233
Haspelmath, Martin, 2, 6, 8, 91, 92,
107, 111, 126, 145, 178, 189
Heim, Irene, 178
Heine, Bernd, 228, 229
Herbeck, Peter, 4
Hill, Virginia, 20
Himmelmann, Nikolaus P., 233
Hofmann, Johann Baptist, 193, 194
Holm, John, 89
Holmberg, Anders, 100
Hong, Jangman, 247
Hopper, Paul J., 97, 165
Howe, Chad, 9, 221–230, 243
Hundertmark-Santos Martins,
Maria Teresa, 225

Iatridou, Sabine, 5
Ihsane, Tabea, 14, 38, 40, 41
Ilari, Rodolfo, 225
Inverno, Liliana Cristina Coragem,
119

Jayeze, Jacques, 67
Johnson, Valen E., 243
Jonckbloet, W. J. A., 210
Jones, Michael Allan, 178, 181, 182,
187, 211

Kailuweit, Rolf, 46–48
Kaufman, Terrence, 120
Kellert, Olga, 2, 8, 48, 49, 52, 59, 61,
65, 75–78, 160, 163, 165–167,
169, 171
Klein, Uli, 40
Kleparski, Grzegorz A., 72
Koch, Peter, 1, 47, 48, 55, 56, 121, 224
Kouwenberg, Silvia, 112, 113
Kuryłowicz, Jerzy, 165

Kyala, Miguel Barroso, 115, 117
Kytö, Merja, 232

Laca, Brenda, 226
Laman, Karl E., 115–117
Lamiroy, Béatrice, 3
Lapesa, Rafael, 168
Lasersohn, Peter, 226
Leal, António, 225
Lehmann, Christian, 165, 170
Lenci, Alessandro, 226
Leonetti, Manuel, 30, 239
Lipski, John M., 88, 103
Loi Corvetto, Ines, 208
Lommatzsch, Erhard, 145
Lopes, Célia Regina dos Santos, 100
Lozza, Arturo Marcos, 65
Lyons, Christopher, 1

Maglia, Graciela, 92–96, 98, 99, 101
Maho, Jouni, 96
Malkiel, Yakov, 195
Maniacky, Jacky, 118
Marchetti, Pablo, 64
Mardale, Alexandru, 20
Maribel, Romero, 163
Maurer, Philippe, 112, 113
McEnery, Anthony Mark, 233
McWhorter, John M., 104
Medina, Enrique, 64, 79
Meeuwis, Michael, 115
Menéndez Pidal, Ramón, 145
Menéndez-Benito, Paula, 2, 48, 67,
142
Mensching, Guido, 178, 179, 211
Meo Zilio, Giovanni, 46, 47
Meyer-Lübke, Wilhelm, 145, 178, 194
Michaelis, Laura A., 226
Michaelis, Susanne Maria, 90

Name index

- Milsark, Gary L., 19
Mingas, Amélia A., 120
Molsing, Karina Veronica, 225
Monaci, Ernesto, 185
Moñino, Yves, 90, 92–96, 98, 99, 101
Moseley, Christopher, 179
Mufwene, Salikoko, 107, 132

Navarrete, María Cristina, 132
Nedelcu, Isabela, 16–18, 25, 36, 38
Nespoli, Juana Barros, 225
Neumann-Holzschuh, Ingrid, 129
Norberg, Dag, 198
Novaes, Celso Vieira, 225

Oesterreicher, Wulf, 47, 48, 55, 56, 224
Olbertz, Hella, 225–227
Oliveira, Fátima, 225

Pagotto, Emílio, 249
Palomo, José R., 145
Pană Dindelegan, Gabriela, 15
Pato, Enrique, 146
Pérez-Saldanya, Manuel, 6, 8
Perini, Mário, 225
Piccitto, Giorgio, 46, 54
Pires de Oliveira, Roberta, 4
Poletto, Cecilia, 14, 189, 190
Poplack, Shana, 231
Port, Angelika, 2, 48, 63, 65, 66
Posio, Pekka, 4
Pozas-Loyo, Julia, 6, 8, 141, 143, 145, 146, 148, 150, 154
Pugliese, Osvaldo, 65
Pustejovsky, James, 226

Quer, Josep, 65, 66

Ramat, Paolo, 186, 189

Real Academia Española, 46, 50
Remberger, Eva-Maria, 179
Rindler-Schjerve, Rosita, 179
Rivero, María Luisa, 142, 143, 145, 166
Rizzo Salierno, Melisa, 52
Roberts, Ian, 75, 97
Roche, Jörg, 118
Rohlf, Gerhard, 185, 193, 195, 197, 198
Romalo, Valeria Guţu, 16
Romano, Roberto, 55
Rosemeyer, Malte, 4, 231, 232, 241, 243, 245, 246, 251
Rosselli, Carlos Patiño, 92, 94, 99
Roussou, Anna, 75
Royer, Justin, 69

Saba, Agostino, 201
Sag, Ivan A., 2
Sanna, Antonio, 208
Santos, Diana, 225
Schirakowski, Barbara, 4
Schmidt-Radefeldt, Jürgen, 119
Schmitt, Christina, 225
Schneider, Edgar, 129
Schuchardt, Hugo, 103
Schwegler, Armin, 89, 90, 93–96, 100, 102, 124
Schwenter, Scott A., 124
Seiler, Guido, 75
Selinker, Larry, 104, 118
Silberstein, Enrique, 55
Solmi, Arrigo, 195
Sophocles, Evangelinus Apostolides, 195
Stan, Camelia, 15
Stark, Elisabeth, 16, 20, 23, 26, 38, 48, 75, 143
Stein, Dieter, 75

Stowell, Tim, 178
Szabolcsi, Anna, 178
Szmrecsanyi, Benedikt, 232

Tasmowski, Liliane, 38
Teruggi, M., 46, 58
The APiCS Consortium, 126
Thomason, Sarah G., 120, 231
Tobler, Adolf, 145
Tola, Pasquale, 190
Torres Cacoulllos, Rena, 233
Tovena, Lucia, 67
Traugott, Elizabeth C., 75, 165
Tropea, Giovanni, 46, 54

van der Auwera, Johan, 231
van der Wal, Jenneke, 118
Van Geenhoven, Veerle, 5
Vanderschueren, Clara, 4
Varela, Soledad, 4
Verveckken, Katrien Dora, 6, 8
Virdis, Maurizio, 200
Vogeleer, Svetlana, 38
von Heusinger, Klaus, 2, 17–21, 40,
41, 66, 142
von Stechow, Arnim, 226
von Wartburg, Walther, 145

Wagner, Max Leopold, 178, 179, 184,
186, 187, 194, 197, 199
Wall, Albert, 4
Werner, Reinhold, 46, 50
Wilmet, Marc, 38
Winford, Donald, 104, 129
Wolfe, Sam, 187, 191

Zamparelli, Roberto, 75, 152, 161, 162
Zanuttini, Raffaella, 124
Zolli, Paolo, 56

Indefinites in Romance and beyond

Due to their flexibility in interpretation, the use of indefinites and other quantificational expressions is highly variable and subject to dynamic processes of language change. The present volume addresses fundamental linguistic questions about language variation and change in Romance quantificational expressions. It focuses on quantificational expressions in language varieties that have not received much attention in the previous literature, such as Old Sardinian, Argentinian Spanish, Palenquero Creole and Cabindan Portuguese, Catalan, Romanian, and others. The studies included in this volume offer new data on these processes of variation and advance theoretical discussions about language variation and change.