Current Issues in Linguistic Theory 360

Points of Convergence in Romance Linguistics

EDITED BY

Gabriela Alboiu

Ruth King

JOHN BENJAMINS PUBLISHING COMPANY

POINTS OF CONVERGENCE IN ROMANCE LINGUISTICS

CURRENT ISSUES IN LINGUISTIC THEORY

AMSTERDAM STUDIES IN THE THEORY AND HISTORY
OF LINGUISTIC SCIENCE – Series IV

ISSN 0304-0763

General Editor

JOSEPH C. SALMONS
University of Wisconsin–Madison
jsalmons@wisc.edu

Founder & General Editor (1975-2015)

E.F.K. KOERNER
Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin

Current Issues in Linguistic Theory (CILT) is a theory-oriented series which welcomes contributions from scholars who have significant proposals that advance our understanding of language, its structure, its function and especially its historical development. CILT offers an outlet for meaningful contributions to current linguistic debate.

A complete list of titles in this series can be found on *benjamins.com/catalog/cilt*

Volume 360

Gabriela Alboiu and Ruth King (eds.)

Points of Convergence in Romance Linguistics
Papers selected from the 48th Linguistic Symposium on Romance Languages
(LSRL 48), Toronto, 25-28 April 2018

POINTS OF CONVERGENCE IN ROMANCE LINGUISTICS

PAPERS SELECTED FROM
THE 48TH LINGUISTIC SYMPOSIUM
ON ROMANCE LANGUAGES (LSRL 48),
TORONTO, 25-28 APRIL 2018

Edited by

GABRIELA ALBOIU
RUTH KING
York University

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM & PHILADELPHIA

Table of contents

Acknowledgements

We would like to thank the authors, as well as all the participants, abstract reviewers, donors and volunteers who contributed to making the Linguistics Symposium of Romance Languages 48th edition (LSRL 48), held in April 2018 at York University in Canada, a success. Special thanks go to our colleagues around the world who volunteered their time and energy to review (and thus substantially improve) the papers selected for publication in this volume. In alphabetical order, these include: Victor Acedo-Matellan, Peter Avery, Larisa Avram, Laura Bafile, Michael J. Barrie, Isabelle Barriere, Judy Bernstein, Bronwyn Bjorkman, Hélène Blondeau, Laura Colantoni, Cristina M. Cuervo, Ricardo Etxepare, Olga Fernández-Soriano, Jon A. Franco Elorza, Ludovico Franco, Elly van Gelderen, Alessandra Giorgi, Virginia Hill, Daniela Isac, Haike Jacobs, Marie Labelle, Maria Lobo, Ruth Lopes, Eric Mathieu, Marcello Modesto dos Santos, Telmo Moia, Alexandru Nicolae, Jairo Nunes, Francisco Ordóñez, Ivan Ortega-Sanots, Anna Pineda, Mihaela Pirvulescu, Lori Repetto, Yves Roberge, Marcos J. Rohena-Madrazo, Johan Rooryck, Andres Saab, Liliana Sanchez, Cristina Schmitt, Michelle Sheehan, Miquel Simonet, Petra Sleeman, Jane Smith, Danielle Thomas, Michelle Troberg, Evani Viotti, and James Walker.

Last but not least, we would like to thank John Benjamins Publishing Company, and the CILT series in particular, for their continued support, guidance and patience, as well as for giving the LSRL one last chance to have some of its papers see the light of day.

Introduction

Gabriela Alboiu and Ruth King
York University

This volume contains a selection of significantly revised, peer reviewed papers presented at the 48th Linguistic Symposium on Romance Languages held at York University in Toronto in 2018.

The fifteen papers selected for the volume provide rich data for a variety of Romance languages past and present centring around the theme of the conference, namely 'Points of Convergence in Romance Linguistics'. The volume, in the spirit of the conference, aims at *bridging* among the various areas of linguistics (e.g., morphology, syntax, semantics, phonology, sociolinguistics, first and second language acquisition) in relationship to both synchronic and diachronic research of single or comparative Romance languages.

The diachronic data come from Old Iberian Romance (Corr) and from Latin, Old Spanish, Old Portuguese and West-Iberian Medieval (Gibert-Sotelo). Several present-day languages are represented: Brazilian Portuguese (Lima & Oliveira; Doner & Bilgin), Catalan (Bembridge & Peters), French (Authier & Reed; Doner & Bilgin; Vogh), Picard (Auger), Romanian (Irimia; Isac), and Spanish (Doner & Bilgin; Gibert-Sotelo; Vázquez-Lozares; MacDonald & Vázquez-Lozares; Tetzloff; Jiménez-Fernández & Tubino-Blanco; Bembridge & Peters). Several papers are comparative in nature: Romanian and Eastern (Isac) or Western (Irimia) Romance; Basque and Spanish (Vázquez-Lozares); Catalan, Peninsular and Latin American Spanish (Bembridge & Peters); Latin and several Romance languages (Gibert-Sotelo); and General Spanish, Brazilian Portuguese, French and Dominican Spanish (Doner & Bilgin).

The bulk of the papers deal with morphosyntax and build largely but not exclusively on recent generative approaches (including current approaches to Minimalism, Cartography, Distributed Morphology, and Nanosyntax) as well as alternative models such as Construction Grammar. A smaller number of papers involve phonology or semantics, and diachronic syntax, the latter stemming from the workshop on 'Romance Diachrony at the Interfaces' held as part of the LSRL 48 symposium.

The papers engage with both longstanding theoretical debates (for citations, see the relevant chapter) such as the status of pronominal clitics as arguments or agreement markers (Auger) and the relationship between the EPP and null subject (Doner & Bilgin). They also engage with theoretical puzzles for specific Romance languages such as the characterization of Spanish stress assignment (Tetzloff), the sequencing of the Spanish *se* clitics (MacDonald & Vázquez-Lozares), the semantic interpretation of bare singulars in Brazilian Portuguese (Lima & Oliveira), cartographic mapping and speech acts in relationship to Spanish *nosotros* (Jiménez-Fernández & Tubino-Blanco), and differential object marking in Romanian (Irimia). The analysis of French-English codeswitching in a minority francophone community is shown to benefit from a semantico-pragmatic discourse analysis approach (Vogh). Other papers use new theoretical tools to account for grammatical and register distinctions in the Spanish pronominal system (Bembridge & Peters) and the relationship between null determiners and prepositions in Eastern Romance (Isac). Another paper applies the relatively recent hypothesis concerning silent elements in the syntax to particular constructions in French (Authier & Reed). Last but not least, language acquisition is represented in a discussion of recursive nominal modification in Spanish (Pérez-Leroux).

1.   The organization of the volume

The volume is divided into four sections, each denoting different aspects of linguistic convergence: *Interfaces, Bridging issues at the CP-TP-vP levels, Bridging issues at the PP-DP levels*, and *Bridging issues in linguistics*. A brief summary of each section and their respective chapters is presented below.

The first section of the volume, *Interfaces*, includes three papers by invited speakers, each focusing on a specific linguistic interface, in particular, the syntax-phonology interface, the syntax-semantics and language acquisition interface, and morpho-syntax interfacing with micro-variation.

In Chapter 1, Julie Auger argues that the analysis of lexical subject clitics in Picard receives a straightforward explanation once we consider the syntax-phonology interface. The author engages with a longstanding theoretical debate, the status of pronominal clitics as arguments or agreement markers in Romance language varieties. This chapter represents the culmination of a series of prior studies of this phenomenon for Picard, an endangered Gallo-Romance language. Given the amount of attention colloquial French pronominal clitics have received in recent decades, along with the fact that Picard speakers are typically bilingual, Auger carefully lays out the similarities and differences in the behaviour of pronominal clitics spoken in the Vimeu region (located in the French department of Somme), drawing on

diagnostics for argument versus agreement status from the relevant literature. She presents strong evidence for agreement status for Vimeu Picard, based on vowel epenthesis, bringing together arguments from both morphosyntax and phonology. The analysis also provides empirical support for a Clitic Group level of representation between the Phonological Word and the Phonological Phrase, thus contributing to another longstanding theoretical debate. Auger thus presents a clear case for what may be learned from research on endangered languages which, she concludes, urgently need further such research given their tenuous status as living languages.

In Chapter 2, Ana Teresa Pérez-Leroux interfaces syntax-semantics with language acquisition. The author addresses a topic well anchored in recent studies: the acquisition of recursive nominal modification by monolingual and bilingual Spanish speakers. The goal, however, is a more challenging one as the author aims at comparing two "different notions of evidence": evidence "that settles theoretical debates" and "plausible learnability evidence". The acquisition data in the chapter offers a comprehensive picture of the early use of recursive nominal modifiers in both monolingual and bilingual contexts (from an impressive number of participants), as well as of the input received by Spanish-speaking children. It is shown that despite the low frequency of PP modifiers in Spanish and various usage restrictions, recursive embedding of PP modifiers can exist for speakers who use those PPs productively, at first level of embedding. In line with requirements stated at the outset of the minimalist program in Chomsky (1995) and Hauser, Chomsky & Fitch (2002), this confirms the non-trivial conclusion that children have "a rich implicit understanding of abstract structure, and work with principles of computational efficiency in the acquisition process".

In Chapter 3, Daniela Isac offers an analysis of micro-variation in Eastern Romance with respect to the expression of definite articles in modified versus unmodified objects of prepositions. The general pattern is that definite articles can be null in unmodified objects of P and must be expressed overtly with modified objects of P (where 'modified' can mean with a PP, an AP, or a relative clause). The micro-variation concerns the issues of where in the DP the definite articles are lexicalized (i.e., on the N, on the A, or on both), as well as whether or not an unmodified object allows or requires a definite article. The analysis adopts notions of (un)interpretable and (un)valued features (Pesetsky & Torrego 2007), the syntactic position/role of prepositions (Pesetsky & Torrego 2004), and builds on work by Matushansky (2006) about head movement (of P) to specifier (of DP) and the author's proposed subsequent m-merge (P with a definite D). Whether or not a definite article is spelled out depends on whether the m-merge occurs with a head bearing a [def] feature. The microvariation observed across Eastern Romance is accounted for by the feature content of the Num head, which varies cross-linguistically, and the "domain of application of the m-merge and spell-out rules".

The second section, *Bridging issues at the CP-TP-vP levels*, includes papers focused on issues involving phenomena residing in various domains of the clausal spine: within the vP (Chapters 4, 5), within the TP-vP interface (Chapters 6, 7) or at the CP-TP interface (Chapter 8).

In Chapter 4, Monica Alexandrina Irimia sets forth to reconcile the observed structural accusative syntax of Differential Object Marking (DOM) with its oblique appearance. The data focus mainly on Romanian but Italian dialects are also discussed. The chapter first evaluates current theoretical proposals on the nature of DOM. In particular, it addresses the 'oblique syntax' model of Manzini & Franco (2016) and the morphological accounts for DOM=OBL proposed by Keine & Müller (2008) and Bárany (2018), among others. Various diagnostics, which include passivization and in situ past participle agreement with a DOM internal argument, show that Romance DOM and its micro-variation cannot be fully captured under any of these implementations. Instead, Irimia proposes that there is an additional licensing mechanism (some extra feature on little v) that gets activated in the presence of marked DP objects. The author labels this extra feature 'PERSON' and, in the spirit of Lopez (2012), argues for its presence on a syntactic head positioned between v and Voice. This feature does not restrict the spell out of DOM, the latter depending on language specific Case syncretism options. In particular, the morphosyntactic shape of a marked DP object that responds to PERSON depends on whether the ACC Case is contiguous to DAT (which results in DAT syncretism) or to LOC(ative), which results in locative syncretism spelled out as DOM-*pe* in Romanian.

In Chapter 5, Almike Vázquez-Lozares considers the wide availability of null objects in the Spanish of the Basque Country of Spain. In contrast with non-Basque Peninsular Spanish, null objects may occur with indefinite, unspecific antecedents in the former variety. Further, their distribution does not obey the Person Case Constraint, which bans the combination of two clitics unless the second is accusative third person (cf. Bonet 1991 for other Spanish varieties). In line with previous research, Vázquez-Lozares makes the assumption that Basque Country Spanish null objects are instances of *pro*; she proposes that this *pro* is licensed in object position by a D(efiniteness)-feature in v, a feature which has become part of the grammar of this Spanish variety through contact with Basque, a language which allows referential null subjects *and* objects. The deletion analysis the author proposes is informed by Roberts (2010), who argues that like the D-feature on T which allows for null subjects, v also has a D feature which allows for null objects. Finally, the author presents evidence from the literature for referential null objects in several non-contact varieties of Spanish where their presence differs from highly restricted (e.g., Madrid Spanish) to widely available (Rioplatense Spanish). These comparisons lead the author to conclude that transfer from Basque may be seen as accelerating the object agreement cycle (van Gelderen 2011).

In Chapter 6, Julianne Doner and Çağrı Bilgin bring together data from a variety of languages and dialects to challenge the generalization that all languages which have an Extended Projection Principle (EPP) type X will also have null subject language type Y (Alexiadou & Anagnostopoulou 1998). They consider two EPP types, D-on-V EPP languages (e.g., Greek and Italian) and DP EPP languages (e.g., English). The survey presents the relevant details for languages which share the same EPP type (DP EPP) but differ in null subject language (NSL) type: Modern French, which is non- NSL; Brazilian Portuguese, a partially NSL language in that only first-person subjects may be phonetically null; and General Spanish, which is consistently NSL. The authors attend to dialectal variation for a number of languages and to diachronic change for particular varieties, presenting, for example, Toribio's (2000) account of the loss of distinct agreement morphology in Dominican Spanish with the unexpected retention of null-subject and Borges & Pires' (2017) analysis of the change from consistent to partial NSL in Goiás BP, arguably triggered by the loss of [D] feature on T. The authors suggest wide usage of overt pronouns precedes the decline of agreement morphology, not vice versa, as is typically argued. Close inspection of the Dominican Spanish facts shows the EPP type remains constant regardless of whether the subject is null or overt. They conclude by calling for further research on languages with the same NSL type but different EPP types to complete a typology of possible EPP/NSL combinations.

In Chapter 7, Jonathan E. MacDonald and Almike Vázquez-Lozares discuss sequences of impersonal *se* along with inherent *se*, reflexive *se* and aspectual *se* in Spanish. Taking Martins & Nunes' (2017) analysis of similar constructions in European Portuguese as their starting point, the authors explore ungrammatical sequences of *se* and beyond, ultimately arguing that the Spanish data do not support an analysis of Imp$_{se}$ merging as the external argument of the non-finite embedded verb that undergoes subsequent movement to the matrix clause. Instead, they offer an alternative analysis for Spanish in which impersonal *se* (Imp$_{se}$) constructions correspond to a structure in which *se* is merged in T, and there is a null pronoun *pro*$_{se}$ merged in subject position (specifier of v). It is further proposed that Imp$_{se}$ lacks number, and therefore cannot license another *se* (*se* or *si*, as in *sí mismo* expressions). In conclusion, it is argued that Martins & Nunes' account of ungrammatical sequences of *se* for European Portuguese cannot be extended to Spanish.

In Chapter 8, Ángel L. Jiménez-Fernández and Mercedes Tubino-Blanco focus on several aspects of the interpretation of the first-person plural pronoun, *nosotros*, in Spanish. These include clusivity, topicality and focus. The authors examine the referential values of *nosotros* in clauses marked with different topics and foci, together with null versus overt contrasts. It is claimed that these factors are not what is responsible for determining inclusiveness. Rather, the authors propose an account in terms of the projection of a Speech Act Phrase/SAP (Speas & Tenny 2003) and

a Logophoric Center (Bianchi 2003) above the clausal left periphery (CP). It is these domains which determine valuation of the [Addressee] feature on *nosotros*. In particular, a Logophoric Operator values the [Addressee] features in Spec of SAP and throughout the lower layer of CP and in TP. The analysis is in line with views that assume the presence of discourse properties in the syntactic derivation (Haegeman & Hill 2013).

The third section, *Bridging issues at the PP-DP levels*, includes four papers focused on issues involving null nominals within PPs (Chapter 9), and overt nominal elements with exploration of features thereof – distributivity (Chapter 10), cardinality and value (Chapter 11), and formality (Chapter 12).

In Chapter 9, Jean-Marc Authier and Lisa A. Reed argue that the properties of French *(ne) … que* exceptives within PP occurrences can be explained by adopting current analyses which assume the presence of a silent n-word (O'Neill 2011; Homer 2015). The chapter contributes to the emerging literature on syntactic silent elements (SEs), argued to be semantically recoverable from their phonologically overt counterparts – a crucial assumption for learnability (Her & Tsai 2015). Authier & Reed revisit two SEs argued to be present in the French exceptive constructions *(ne)…que*: silent *rien* "nothing" and silent *autre* "other" (O'Neill 2011; Homer 2015). On the basis of data from both colloquial and formal French, Authier & Reed argue that while *rien* is indeed the semantic equivalent of its phonologically overt counterpart in the relevant construction, *autre* is not. In the latter case, they argue that the SE is actually *plus* "more", suggested in part by comparisons with Spanish and by tests for semantic equivalence and hence recoverability.

In Chapter 10, Alice Corr provides novel data for the little-studied phenomenon of reduplicated numerals in Old Ibero-Romance, more specifically Old Spanish, Old Portuguese and West-Iberian Medieval Latin. While the earliest textual evidence is from the 10th–15th centuries, the bulk of the data are drawn from 13th–14th century documents. The construction, which the author labels distance distributivity with numerical reduplication (DDNR), is both typologically rare (it is not found elsewhere in Latin or Romance) and typologically unexpected, given that reduplication is unproductive elsewhere in (Ibero-)Romance. In this first theoretical account of the phenomenon, the author argues that DDNRs behave like double object constructions and argues for an analysis with two Applicative projections, one high and one low (in the spirit of Pylkkänen 2008) over a small clause analysis. She situates DDNR constructions within a wider typology of Ibero-Romance clausal constructions and shows the utility of a constructivist approach to encoding distributivity in the nominal functional structure, an approach in keeping with previous analyses (Beghelli & Stowell 1997; Stowell 2013) of the English binomial *each*.

In Chapter 11, Suzi Lima and Cristiane Oliveira investigate experimentally the effects of two semantic features, cardinality and value, on the interpretation of

bare singulars in argument positions in Brazilian Portuguese, the second of which being a phenomenon not known to occur in other Romance languages. One finding previously reported in the literature is confirmed, in that in neutral contexts a cardinal response is favoured in questions including a bare singular or plural but disfavoured in questions including a mass noun. The authors go on to investigate the effects of cardinality and value in two additional experiments, one in which the contexts were manipulated such that value was more relevant in one context than in another and one in which a priming task manipulated these same two features. Overall, cardinality is found to be more relevant grammatically than value, evidence for a distinction between languages like BP and languages like English (Grimm & Levin 2012), where both features are at play, in the interpretation of mass nouns.

In Chapter 12, Gavin Antonio Bembridge and Andrew Peters propose a Distributed Morphology account of formality in second-person pronouns across varieties of Spanish and Catalan. The authors encode formality as a form of 'social distance' using the syntactic head χ from Harbour (2016), as adapted to pronominal systems in Bjorkman et al. (2019). More specifically, it is argued that Spanish and Catalan second-person pronouns can encode both personal and spatial deixis. The presence of the locative χ feature positions the addressee close to or far from the characteristic space of the author. Variation in second-person pronouns results from differences in the way χ combines with person features, as well as details of each language's specific morphological resources (i.e., Vocabulary Items).

The fourth and last section, *Bridging issues in linguistics*, includes three papers focused on bridging issues concerning diachrony and Nanosyntax (Chapter 13), discourse-analysis and variation (Chapter 14), and phonology (Chapter 15).

In Chapter 13, Elisabeth Gibert-Sotelo investigates the differences which are found between the negative prefix *in-* in Latin and in Romance, arguing for reanalysis from an adjunct to a categorizing affix, an example of a partial negative cycle (cf. van Gelderen 2011). The author outlines the differences between the Latin prefix, which behaves as a morphological adjunct that may attach to scalar predicative roots denoting (in)alienable possession or inherent properties, and the Romance prefix (as exemplified primarily with Spanish and Catalan data), which may only attach to predicate adjectives. From the perspective of the relatively recent Nanosyntax model (Starke 2009) in which the analysis is embedded, this case of reanalysis may be characterized as changing the size of the lexically stored tree through the addition of a particular grammatical feature, an adjectivizing head. Gibert-Sotelo argues for the usefulness of such an approach for capturing both synchronic linguistic diversity and, in the present case, diachronic change.

In Chapter 14, Kendall Vogh focuses on code-mixing in a corpus of semi-structured interviews with elderly Franco-Americans living in rural Maine, where francophones make up a small and diminishing proportion of the population. Despite a

long history of little institutional support or any social or economic benefits for use of French, these individuals are fluent French-English bilinguals whose skills in the minority language are primarily oral. Their linguistic repertoires contain substantial amounts of French-English codeswitching, usage which the author argues to function as an important semantico-pragmatic resource for in-group conversations. In this chapter, Vogh is primarily concerned with four frequently occurring lexical items – *oui, ouais, yes* and *yeah* – which arguably constitute a linguistic variable since they may all be used for six semantico-pragmatic categories of meaning-in-use, i.e., affirmatory response, alignment, emphatic expression, continuer, subject change, and repair. However, differences do emerge in that the French units are preferred for those meanings and functions involving (inter)subjectivity, such as for emphatic expression and other-oriented repair. Vogh suggests that this finding is not unexpected, given that the primary function of French in this community involves the re/production of a sense of group belonging and social cohesion. Like Auger's contribution for Picard, Vogh shows the importance of research on endangered languages, not only for the historical record, but for what such research can contribute to theory construction across a variety of subfields.

In Chapter 15, Katie Tetzloff proposes an analysis of Spanish nominal stress couched within the stratal OT approach. The author proceeds by first engaging with previous analyses of Spanish stress, especially Baković (2016). The central tenet is that regular patterns can be accounted for by assuming that the stem is quantity-sensitive, and exceptions can be modeled using lexically-indexed constraints. In this way, it is possible to account for exceptional patterns without generating ungrammatical patterns. The latter distinction is an important one and a central contribution of this chapter. The chapter concludes this is so "due to the unique proposal that Spanish stress assignment is sensitive to the weight of the morphologically-derived word-stem".

In sum, this volume provides an overview of recent research on idiosyncratic topics in Romance linguistics, presenting novel data, engaging current theoretical issues and employing a number of methodological advances. The chapters shed new light on a variety of longstanding debates in the field and also engage with novel ones. Taken together, they showcase the richness of current research and suggest several intriguing directions for future research.

Interfaces

Picard subject clitics

An analysis at the interface of syntax, phonology, and prosody

Julie Auger

Université de Montréal

This paper argues that the analysis of Picard subject clitics and lexical subjects greatly benefits from considering syntax, phonology, and prosody. Specifically, the status of pronominal clitics as arguments or agreement markers must be determined on the basis of syntactic criteria, while the determination of their status as clitics or affixes necessitates phonological arguments. My analysis shows that Picard subject clitics behave like agreement markers but that they retain the status of clitics, thus providing evidence for the Clitic Group as proposed by Nespor & Vogel (1986). I also provide evidence that other grammatical monosyllables do not create clitic groups.

Keywords: subject clitics, subject doubling, clitic group, vowel epenthesis, prosodic hierarchy, agreement marking, grammaticalization, colloquial French, dislocation, stylistic inversion

1.   Introduction

Even though numerous articles and monographs are devoted to analyses of Romance pronominal clitics, many central issues remain hotly debated (Heap, Oliviéri & Palasis 2017). Are pronominal clitics generated in syntax or in morphology (e.g., Gaglia & Schwarze 2015)? Are they syntactic arguments or agreement markers (e.g., Fuβ 2005; Culbertson 2010)? What are the linguistic factors that favor their grammaticalization (e.g., Ashby 1977; Martineau & Mougeon 2003)? Is there any evidence for a clitic group level in the prosodic hierarchy (Nespor & Vogel 1986; Vogel 2009)? French subject clitics, in particular, have been the focus of much debate. Their propensity to co-occur with lexical subjects (e.g., Nadasdi 1995 for Franco-Ontarian; Beaulieu & Balcom 1998 for the Acadian variety of northeastern New Brunswick; Fonseca-Greber 2000 for the Swiss variety; Coveney 2003 for Picardie French; and Auger & Villeneuve 2010 for Québec French) has led many

scholars to propose that colloquial French's subject clitics have been reanalyzed as agreement markers. However, other studies have shown that such grammaticalization is not universal among French varieties. This is the case, for example, for the variety of Acadian French spoken in Newfoundland (King & Nadasdi 1997) and the Belgian data analyzed by De Cat (2007). One element that has contributed to the debate concerning the status of French subject clitics is the fact that the frequency with which subject clitics co-occur with lexical subjects varies across varieties, with frequencies ranging from 21% in Tours (Ashby 1980), 24.4% in Picardie (Coveney 2005), and 27% in Ontario (Nadasdi 2000) to 55% in Montréal (Sankoff 1982). To my knowledge, the only studies reporting near-categorical subject doubling in French are Campion's (1984) analysis of adolescents in Villejuif, a suburb of Paris, where 96.4% of the subjects were doubled, and Palasis' (2013) study of preschool children in southern France, which shows that subject doubling starts out as almost categorical and decreases as children become aware of the social stigma attached to this syntactic construction. In addition to considerable variation across communities, research also reveals that the frequency of subject doubling greatly varies across speakers. For example, Auger & Villeneuve (2010: 75) report frequencies of subject doubling ranging from 20% to 79% in the Saguenay (Québec) variety of French. Other factors that can explain the persistence of the debate concerning French subject clitics include the small number of existing empirical studies and the strong pressure that prescriptive forces continue to exert not only on speakers, but also on analysts.

What would contemporary colloquial French look like if it had not been the object of so much interference on the part of an intellectual elite who has expressed strong opinions about what constitutes "good French", directly intervened in attempting to steer its evolution, and influenced the role played by parents and teachers in their children's linguistic acquisition? Descriptions and prescriptive comments from past centuries reveal that some features of contemporary colloquial French have been attested and criticized for a long time. For example, Vaugelas' *Remarques sur la langue françoise*, published in 1647, documents the use of *je vas* "I go" instead of *je vais*, haplology for *le/la lui* 'him/her to him/her' realized as *lui*, and numerous other forms that persist to this day. More importantly for us, Oudin (1632: 82) makes it clear that subject doubling was already a frequent feature in 17th century French when he writes: "On ne met point de pronom personnel apres un substantif, pour servir à un mesme sujet: par exemple on ne dit jamais, *Monsieur il a*, mais, *Monsieur a dit*".[1] Given that subject doubling has been part of the French language for at least 500 years, we might expect that this construction would have

---

1.   "One doesn't put a personal pronoun after a noun in order to refer to the same subject: for example one never says *Sir he has* but *Sir has said*".

become categorical by now. This is obviously not the case, as prescriptive grammars of French continue to recognize only non-doubled subjects as standard.

While it is impossible to turn back the clock to determine how French might have developed in a less prescriptive society, different approaches can help us gain a better sense of what contemporary vernacular French is or could have been. One of them involves the analysis of young children who have not yet been subjected to strong normative pressure. Palasis (2013, 2015) shows that in the data from the youngest children in her corpus (aged 2;3–3;1), lexical subjects are virtually always followed by a pronominal copy, that the standard SVO structure with no pronominal copy gradually emerges as the children learn standard French, and that non-doubled subjects remain marginal in the speech of her oldest children (aged 3;6–4;11). Another approach focuses on varieties of French spoken in locations and communities in which schooling and normative pressure have been weak and where one can expect to observe a vernacular less affected by standard French. Such studies reveal important differences across varieties: for example, while subject clitics show no evidence of grammaticalization in the Newfoundland variety of Acadian French analyzed by King & Nadasdi (1997), the quasi-categorical subject doubling observed in the speech of teenagers in the Parisian suburb of Villejuif by Campion (1984: 219) suggests that their reanalysis as agreement markers is complete in this variety. A third option consists of examining other Oïl varieties that are closely related to French. Since varieties such as Norman, Poitevin, and Burgundian do not have official recognition and have been used mostly in daily oral communication and very little in "serious" written form, we can hypothesize that their respective grammars have been freer to follow their normal course and thus can give us a realistic sense of what colloquial French would resemble with respect to features that they share. This approach is the one that we pursue in this paper.

Picard is an Oïl variety that is spoken in northern France and in southern Belgium. Even though considerable variation characterizes such a large geographical area, there is one feature that is mentioned in virtually every description of Picard and that distinguishes it from the neighboring varieties of Walloon and Norman: subject doubling (Dawson 2010: 24 and Edmont 1897: 10 for the Pas-de-Calais region; Ledieu 1909: 42, Hrkal 1910: 262, and Debrie 1974: 18 for Amiénois Picard; Cochet 1933: 36 and Dauby 1979: 43 for the Nord *département*; and Vasseur 1996: 61 for Vimeu). For instance, Vasseur writes "Les substantifs ne s'emploient jamais seuls comme sujets; on les fait suivre d'un pronom personnel qui précède immédiatement le verbe et qui s'accorde en nombre et en personne avec le sujet",[2] and he provides the following examples:

---

2. "Nouns are never used alone as subjects; they are followed by a personal pronoun that immediately precedes the verb and that agrees in number and person with the subject".

(1)   Vasseur's examples of subject doubling[3]

   a.   *Ch' beudet i   minge*
        the donkey he eat
        "The donkey eats"

   b.   *Chov vaque ad- donne du   lait*
        the   cow she give of-the milk
        "The cow gives milk"

   c.   *Chés vieux i   boét't*
        the   calves they drink.3PL
        "The calves are drinking"

   d.   *Chés hirondélles i   s'   in   vont*
        the   swallows   they SELF of-there go
        "The swallows are leaving"

Subject doubling in Picard differs from its French counterpart in at least two important respects: the fact that it affects all subject types and that it is nearly categorical. In (2), we can see that, in addition to lexical DPs, proper names, and strong pronouns, bare quantifiers can be doubled by a subject clitic that shares the same person, number, and gender features. Given that bare quantifiers are not allowed in dislocated positions (Rizzi 1986), the fact that those subjects can be doubled is interpreted by many scholars as strong evidence that languages in which such doubling is observed feature true subject doubling rather than dislocation (Brandi & Cordin 1989; Poletto 2000).

(2)   Subject doubling in Picard

   a.   *Min grand-pére il   étoait coér in route à lacher ses solés*
        my grandfather he was   still in road to lace   his shoes
        "My grandfather was still lacing his shoes"   (Chl'autocar 17)

   b.   *Fonse i   n'   étoait point[4] lo*   (Chl'autocar 18)
        Fonse he NEG was   not   there
        "Fonse wasn't there"

   c.   *si élle al   prind ch' car éch mérquédi*   (Chl'autocar 40)
        if her she take the bus the Wednesday
        "si SHE takes the bus on Wednesday"

---

3.   Vasseur's examples are provided in phonetic transcription. I provide my own transliterations based on the orthographic convention that he and other Picardisants have developed for Picard.

4.   Contrary to modern French, which has preserved only one post-verbal negation, *pas*, Picard varieties typically possess two of them (Dagnac 2015). In Vimeu Picard, *point* is the general negator and *mie* has a more limited distribution, being associated with expressive uses and contradiction (Burnett & Auger 2016).

    d.   *Parsonne i n'   poroait mie vnir ll'   értcheure*   (Chl'autocar 40)
        nobody he NEG could   not come him get again
        "Nobody would be able to come and get him"

Auger (2003a) confirms that, in her written data, subject doubling affects all overt subjects, and Villeneuve & Auger (2013) found subject doubling to be near-categorical in the speech of four speakers. However, given Coveney's (2005: 103) observation that subject doubling "has particular significance as a badge of Picard identity", the possibility arises that it might not be a real feature of the grammatical competence of Picard speakers but rather an element that they consciously insert into their Picard clauses in order to make them 'more Picard'. In the current context of revitalization of a language that is spoken by fewer and fewer individuals, one can expect to find considerable differences in the quality of the language used in oral and written productions. While some original texts and translations faithfully mirror the Picard varieties spoken by the latest generations of fluent speakers, others are replete with instances of stereotypical features and simplified use of Picard shibboleths (cf., e.g., the focus on [k] and [ʃ] instead of French [ʃ] and [s], respectively, in the *Bienvenue chez les Ch'tis* movie; Dawson 2008) and even erroneous forms, such as the use of the masculine instead of feminine determiner in the title *Ch' bièle provinch'* for a translation of a Lucky Luke book into the Chti variety of Picard spoken in the Nord and Pas-de-Calais departments (Landrecies 2006: 72). Consequently, we need to determine whether subject doubling is part of the grammar of Picard or simply an inserted stereotypical feature. In order to do this, we focus on the variety of Picard of the Vimeu region in the Somme *département* in France to answer three different questions. First, we seek to establish whether subject clitics truly behave like agreement markers and thus appear in all constructions in which a tensed verb is expected to agree with its subject. Second, we ask whether any independent evidence is available for determining whether the doubled subjects are syntactic subjects or dislocated phrases. Third, we examine the subject clitics themselves in order to determine whether they still behave like phonological clitics or whether they have become lexical affixes.

2.   Subject clitics: Syntactic subjects or agreement markers?

Sentences such as those in (2), in which the subject appears to be expressed twice, raise questions concerning their structure. Given the widely accepted view that verbs can assign nominative case to only one subject, the possibility that both the DP/strong pronoun and the clitic are subjects is ruled out. This leaves the analyst with a choice between two possible structures: either the clitic is the syntactic

subject and the DP/strong pronoun is a dislocated phrase that is set apart from the core sentence to achieve some pragmatic effect (e.g., emphasis, contrast, introduction of a new topic; cf. Barnes 1985 and Ashby 1988), or the DP/strong pronoun fulfills the subject function and the pronominal clitic has been reanalyzed as a pre-verbal agreement marker. While the former analysis is typically adopted for standard French, many researchers have argued that the latter best describes colloquial French (e.g., Roberge 1990; Auger 1994; Zribi-Hertz 1994 and Legendre et al. 2010; but see King & Nadasdi 1997 for discussion of a variety in which subject clitics have not been reanalyzed as agreement markers and De Cat 2007 for a rejection of this analysis for colloquial French). Those who favor the agreement-marking analysis base their position on the fact that in these varieties of French, subject clitics occur in all contexts in which a verb is expected to agree with its subject. For instance, subject clitics co-occur with bare-quantifier subjects in (3a), they are repeated on each conjunct in a VP-conjunction structure in (3b), and they occur in subject relative clauses in (3c), and in inverted constructions in (3d).[5]

(3)   Subject clitics in Québec Colloquial French
   a.   *en campagne, quand quelqu'un il   dansait…*   (Auger 1994: 97)
        in countryside when   someone he danced
        "In the countryside, when someone danced…"
   b.   *il a   laissé ça   pis il a   rentré   à Northern*   (Auger 1994: 77)
        he has left   that and he has entered at Northern
        "he quit that [job] and went to work at Northern"
   c.   *J' étais pas une personne que   j' avais beaucoup d' amis*
        I was   not a   person   COMP I had   many   of friends
        "I wasn't someone who had many friends"   (Auger 1994: 77)
   d.   *Je me demande où   ce   qu'   elle est sa   maison*
        I me ask   where that COMP she is her house
        "I wonder where her house is"

The criteria that support an agreement-marking analysis for colloquial French subject clitics provide support for the same analysis in Picard. In this language, all subjects are doubled, including bare quantifiers, as seen in (2d). Subject clitics are present in subject-verb inversion constructions as in (4); they are repeated on each verb in a VP-conjunction as in (5); and they occur in subject relative clauses as in (6). In subject *wh*-questions, a default third person masculine singular marker is used, as illustrated in (7).

---

5.   The data in Auger (1994) come from the Sankoff/Cedergren corpus of Montréal French collected in 1971.

(4) a. *j' édmanne à quiqu'un doù   qu'   i   reste Fonse*   (Chl'autocar 35)
         I   ask   to someone where COMP he live   Fonse
         "I ask someone where Fonse lives"

     b. *qu'   a   s'   a   dépétchè d' dire inne féme*   (Chl'autocar 19)
         COMP she SELF has hurried of say a   woman
         "that a woman rushed to say"

(5) *al   a   rougi   pi al a   tornè   s'   téte*   (Chl'autocar 50)
     she has blushed and she has turned her head
     "she blushed and she turned her head"

(6) a. *inne dame qu'   a   mé zz'   éroait   pétète acatès*
         a   lady COMP she me them would.have maybe bought
         "a woman who might have bought them from me"   (Chl'autocar 20)

     b. *des   gins   qu'   il   étoait't din chés camps*   (Chl'autocar 21)
         of-the people COMP they were   in   the fields
         "people who were in the fields"

(7) *tchèche qu'   il   éroait   peu   prévoér tout o?*
     who   COMP he would.have been-able-to predict all   that
     "who could have predicted all that?"   (Chl'autocar 28)

The regular presence of subject clitics whenever a verb agrees with its subject provides strong evidence that the subject doubling observed in Picard results from the reanalysis of subject clitics as agreement markers and that Picard is a null-subject language, like many North Italian dialects (cf., e.g., Brandi & Cordin 1989 and Poletto 2000). While the standardization process that is underway may have contributed to make this reanalysis progress a little faster through editorial policies that insert subject clitics in clauses that lacked them (Auger 2003b, 2018), there is no doubt that the completed grammaticalization marks the culmination of a process that was already nearing completion and that subject doubling is a "badge of Picard identity" (Coveney 2005: 103) that constitutes an integral part of the grammar of the language.

3.   Lexical subjects: Syntactic subjects or dislocated phrases?

Section 2 provided evidence that the subject clitics of Picard have been reanalyzed as agreement markers, which opens the possibility that doubled lexical DPs and strong pronouns are syntactic subjects rather than dislocated phrases. The fact that subject doubling is categorical or near-categorical in Picard further supports this analysis, as it is clear that this construction cannot serve to convey special pragmatic functions of contrast and emphasis that are typically associated with dislocation. In this section, I provide phonological evidence that most doubled subjects, whether

in preverbal or postverbal position, do not constitute their own intonational phrases in the sense of Nespor & Vogel (1986) but rather belong to the same intonational phrase as their verbal host and its other arguments.

3.1   Vowel epenthesis in Picard

Strategies for dealing with sequences of consonants that cannot be syllabified include deletion, assimilation, and vowel epenthesis. In the Vimeu variety of Picard, deletion is restricted to a few suffixes (e.g., -iste, as in jornalisse "journalist"), and vowel epenthesis is used whenever a sequence of three or more consonants needs to be syllabified. What makes epenthesis an excellent diagnostic for prosodic structure is the fact that the modalities of its application differ based on the type of structure in which it occurs.

The syllable structure of Picard closely resembles that of French. Specifically, many branching onsets are allowed, as long as sonority rises, the sonority distance between them is sufficient, and the two consonants do not share the same place of articulation (e.g., [pr] and [mj] are possible onsets, but not *[rk], *[rj], or *[ʃl]). Auger & Hendrickson (2011) provide evidence that Picard allows branching codas if they involve a liquid followed by a less sonorant consonant (e.g., [rl] or [lm], but not *[pt], *[nʃ], or *[st]). When morphology or syntax creates consonant sequences that do not meet these conditions, Picard inserts its default vowel, [e]. The examples in (8)–(10) illustrate this alternation in three different contexts: word-initial, word-final, and inside clitic groups. In each case, the a. example shows that when only two consonants are present, they are syllabified into separate syllables, and the b. example shows the insertion of [e] in the presence of three consonants.

(8) Word-initial epenthesis
    a. *Chl' autocar il a cminchè* [akmɛ̃ʃɛ] *à tranner*   (Chl'autocar 19)
        the bus   he has started   to shake
        "The bus started to shake"
    b. *pour écmincher* [purekmɛ̃ʃe] *l' moédeut*   (Chl'autocar 75)
        for   begin   the harvest
        "in order to start the harvest"

(9) Word-final epenthesis
    a. *qu' i cminch't à* [kikmɛ̃ʃta] *rappreucher…*   (Chl'autocar 25)
        COMP they begin 3PL to   approach
        "who begin to approach"
    b. *qu' i cminch't éd* [kikmɛ̃ʃted] *boéne heure*   (Chl'autocar 44)
        COMP they begin 3PL of   good hour
        "that they start early"

(10) Clitic group epenthesis
  a. *I m' sanne* [imsãn] *qu' o sonme d' age*   (Chl'autocar 51)
    it me seem   COMP we are   of age
    "It seems to me that we are the same age"
  b. *J' ém souvarai* [ʒẽmsuvare] *longtemps d' chés jours lo.*
    I me remember   long   of the days there
    "I'll long remember those days"   (Chl'autocar 46)

In all three prosodic contexts, epenthesis allows for sequences of consonants to be fully syllabified. However, its operation differs across prosodic contexts. Auger's (2001: 264) analysis of word-initial epenthesis revealed that it operates categorically inside intonational phrases but variably at the beginning of intonational phrases and utterances. For example, if the juncture of a preposition and a determiner or that of a verb and an adverb creates a sequence of three unsyllabifiable consonants, epenthesis must take place, but if the same juncture contains only two consonants, epenthesis does not take place. As (11) illustrates, if an illicit onset occurs at the beginning of an utterance, epenthesis applies variably. (11) also shows that word-final epenthesis differs from word-initial epenthesis: whereas the latter applies categorically inside intonational phrases, the former applies variably. Auger (2000) shows that the frequency with which word-final epenthesis applies is connected to the size of the prosodic domain in which it occurs: the smaller the prosodic domain, the more frequent epenthesis is. This characteristic of word-final epenthesis is responsible for another difference between word-initial and word-final epenthesis: whereas word-initial epenthesis is variable at the beginning of the utterance, word-final epenthesis is not allowed at the end of the utterance, as seen in (12).

(11) Utterance-initial variation
  a. *Rmérqué bién* [r.mer.ke.bjɛ̃]   (Chl'autocar 38)
    note   well
    "Please note"
  b. *Érmérque bién* [er.merk.bjɛ̃]   (Chl'autocar 38)
    note   well
    "Please note"

(12) *a fzoait juste/*justé* [ʒust]/* [ʒuste]   (Chl'autocar 30)
    it made tight
    "it was tight"

3.2   Subjects and epenthesis

Because epenthesis is sensitive to prosodic boundaries, we can use it to determine whether preverbal and postverbal doubled subjects occur within the same intonational phrase as the verb or in a separate one. If word-initial epenthesis applies categorically at the juncture between the doubled subject and the clitic group that follows it, this supports the analysis of the subject as a syntactic subject that belongs to the same intonational phrase as the verb; if epenthesis applies variably, this supports the dislocation analysis. As for post-verbal subjects, the frequency with which word-final epenthesis applies can help determine whether a major prosodic boundary typical of dislocation intervenes between the VP and the subject.

In the generations of Picard speakers and authors born after 1930, doubling affects all subjects, including quantified ones. However, some older speakers born before 1930 and more conservative speakers born as recently as the late 1950s fail to double bare quantifiers. In such examples, epenthesis applies categorically, such that it obtains after *personne* "nobody" but not *rien* "nothing", as seen in (13). Similarly, at the juncture between a lexical subject or strong pronoun and the following subject clitic, epenthesis generally follows the same pattern, as we can see in (14). Whether they are doubled or not, DP subjects thus behave like syntactic subjects rather than dislocated phrases.

(13)   Preverbal quantified subjects
   a.   *Personne én' veut* [pɛr.sɔ̃.nen.vø] *pu   s'   y   mette.*
          nobody   NEG want   anymore SELF there put
          "Nobody wants to be there anymore"   (Lettes 175)
   b.   *Rien   n'   va* [rjɛ̃.vɑ]   *pu.*   (Lettes 171)
          nothing NEG go   anymore
          "Nothing is going right anymore"

(14)   Doubled preverbal subjects
   a.   *Min corps éch n'   est* [mɛ̃.kɔ.reʃ.ne]   *pu   un ami*
          my   body it   NEG is   anymore a friend
          "My body no longer is a friend"   (Dufrêne 43: 10)
   b.   *mi j' n'   avoais* [miʒ.na.vwɛ] *pu   qu' inne coeuchètte*
          me I NEG had   anymore only one sock
          "me I had only one sock left"   (Chl'autocar 26)

Exceptions to this general pattern are found in contexts that involve a prosodic boundary between the subject and the verb phrase, as seen in (15). Such examples are reminiscent of what we see at similar junctures between a parenthetical clause or the end of a non-restrictive relative clause and the rest of the sentence, as in (16).

This means that clauses containing doubled subjects are ambiguous: the majority of them involve syntactic subjects, but some involve dislocated subjects and null subjects. While many scholars argue that overt subjects in null-subject languages occur either in a dislocated position or in SpecTP (cf. Suñer 2003, among others, for discussion), the phonological evidence in Picard provides evidence that both positions are available.

(15) Dislocated subjects

    a. *Mi, éj n' ai* [mi.eʒ.ne] *rièn   intindu*   (Deglicourt 29: 23)
        me I NEG have   nothing heard
        "Me, I haven't heard anything"

    b. *Ém casaque, ch' n' est* [ka.zak.ʃ+ne] *rien.*   (JVasseur 63: 8)
        my coat   it NEG is   nothing
        "My coat, it's nothing."

(16) Epenthesis and its absence at prosodic boundaries

    a. *Mi, qu' i répond, éj m' ai* [re.pɔ̃.eʒ] *tnu…*   (Rinchétte 2)
        me COMP he reply   I me have   kept
        "Me, he answers, I have spent … "

    b. *écht honme leu, qu' il a tant d' mérite, ch' n' est…*
        that man   there COMP he has such of merit it NEG is
        [me.rit.ʃ+ne]   (Rinchétte 187)
        "that man, who has so much merit, it isn't … "

Vowel epenthesis also provides evidence that some postverbal subjects are VP-internal. Specifically, in clauses that allow for stylistic inversion in French, namely, in *propositions incises* "quotative clauses", in *wh*-questions, whether matrix or embedded, in relative clauses and in subjunctive clauses (Kayne & Pollock 1978), word-initial epenthesis applies categorically when required, as illustrated in (17) and word-final epenthesis is frequently attested, as seen in (18), as is characteristic of phrase-internal positions, but not phrase-final ones.

(17) Postverbal subjects: word-initial epenthesis

    a. *qu' al foait m' grand-mére* [kal.fwɛm.grɑ̃.mer]   (Réderies 17)
        COMP she do   my grandmother
        "says my grandmother"

    b. *qu' i li   dmanne él Diabe* [ki.lin.mɑ̃.nel.djɑb]   (Viu temps 9)
        COMP he to-him ask   the devil
        "asks him the devil"

(18) Postverbal subjects: word-final epenthesis
   a. *quoé qu' i foait'té tous chés gins lo* [ki.fwɛt.te.tu.ʃe.ʒɛ̃.lɔ]
      what COMP they do.3PL all the people there
      "what are they doing all those people" (Chl'autocar 24)
   b. *qu' i dit'té chés ouvrieus* [ki.dit.te.ʃe.zu.vri.ø] (Lettes 619)
      COMP they say.3PL the workers
      "say the workers"

4. Subject clitics: Affixes or clitics?

Agreement markers typically are lexical affixes that cannot be separated from their hosts (e.g., *mangeons* "eat.1PL", *mangerez* "eat.FUT.2PL"). However, as argued by Auger (1994) and Culbertson (2010: 88), the status of subject clitics "as independent words, phonological clitics, or affixes on the one hand, and their FUNCTION as syntactic arguments or agreement markers on the other hand" must be distinguished and determined based on different criteria. Indeed, crosslinguistic data reveal the existence of verbal affixes that are arguments (e.g., the subject agreement affixes of Celtic languages, which cannot co-occur with overt lexical subjects) or of clitics that are agreement markers (e.g., Spanish dative clitics and Bantu subject markers). As stated above, the morphophonological status of the Picard subject clitics must be determined based on phonological criteria. How do subject clitics and their verbal hosts behave with respect to phonological processes that operate differently word-internally and across word boundaries? For Picard, the answer to this question comes from two processes. First, vowel epenthesis provides evidence that pronominal clitics behave differently from non-pronominal grammatical words such as *d* "of/from" and *qu'* "complementizer". Second, the creation of branching attacks distinguishes pronominal clitics from lexical affixes, thus providing evidence for the existence of the Clitic Group that includes pronominal clitics and their hosts as a separate prosodic level, as proposed by Nespor & Vogel (1986) and argued for by Vogel (2009).

4.1 Phonological phrase and clitic group

We saw above that vowel epenthesis provides evidence that most preverbal doubled subjects are syntactic subjects that occur within the same Intonational Phrase as the verb and postverbal doubled subjects that occur in constructions that feature stylistic inversion in Standard French are VP-internal subjects rather than dislocated phrases. In this section, we will see that epenthesis sites distinguish pronominal clitics from other types of words, including non-pronominal monosyllabic.

Word-initial epenthesis follows a strict CeCC template. This can be seen in (19), where three tokens of the preposition *d* "of/from" can be seen. In *chl'éspréss d' Amiens*, no epenthesis is required because both consonants, /s/ and /d/, can be syllabified into different syllables. However, in the first two instances, epenthesis is required due to the presence of illicit consonant sequences. In *vnoéme éd déchénne*, a sequence of three consonants, /mdd/, obtains. Since syllabification operates from right to left in Picard, [e] is inserted before the preposition, resulting in *éd*. In *déchénne dé chl'éspréss*, an unsyllabifiable sequence of four consonants obtains: /ndʃl/. Because /ʃl/ is not a possible branching onset in Picard due to the fact that both consonants share the same place of articulation and because syllabification proceeds from right to left, the epenthetic vowel is inserted after the preposition, producing *dé*.

(19) O vnoéme <u>éd</u> déchénne <u>dé</u> chl' éspréss <u>d'</u> Amiens
      we came   of   get-off   of the express of Amiens
      [ɔ.vnwe.med.de.ʃẽn.deʃ.les.pres.dã.mjẽ]   (Chl'autocar 17)
      "We had just gotten off the express bus from Amiens"

The same apparent variation in epenthesis site can be observed in sequences of pronominal clitics and their verbal host, as shown in (20). However, in this case, both examples contain exactly three consonants, which means that a different explanation must be sought for the different epenthesis sites.

(20) Epenthesis inside clitic groups
      a. *j' ém ramintuve coér bien* [ʒẽm.ra.mẽ.tyv]   (Chl'autocar 55)
          I me remember still well
          "I still remember well"
      b. *j' té   raméne   din tin   poéyi!* [ʃ.te.ra.mẽn]   (Chl'autocar 77)
          I you SG bring.back in   your village
          "I'm taking you back to your village"

Table 1 summarizes the distribution of the CCeC and CeCC patterns in combinations of pronominal clitics and their hosts. It shows that the epenthesis site is determined by the sonority of the first and second consonants: the CCeC patterns is observed when the second consonant is less sonorous than the first or when the two consonants are of equal sonority, while the CeCC pattern occurs when the second consonant is more sonorous than the first (cf. Clements' (1990) sonority scale), that is, in sequences that violate the Syllable Contact Law that disfavors onsets that are more sonorous than the coda that precedes them. This observation led Auger (2003c) to propose that clitic groups have their own epenthesis template, CCeC, but that this template can be modified in order to avoid a violation of the Syllable Contact Law, thus producing the template already familiar for word boundaries: CeCC.

Table 1. Epenthesis patterns within clitic groups*

CCeC	CeCC
ʒteC	ʒemC
nmeC	ʒenC
nteC	ʒelC
nseC	melC
	nelC

* *n* 'NEG' is not a pronoun; however, it is a clitic form that occurs between the subject clitic and the other pronominal clitics and is subject to the same phonological constraints as the other clitics.

Interestingly, the effect exerted by the Syllable Contact Law does not extend to other monosyllabic grammatical words. This is illustrated with the preposition *d* "of/from" in Table 2, which shows that, no matter whether /d/ is followed by equally sonorous consonants such as /t/ and /f/ or by more sonorous ones, such as /n/, /l/, /r/, and /j/, the epenthetic vowel is always inserted between the first and the second consonant, even though its presence between the second and the third would avoid violations of the Syllable Contact Law.

Table 2. Non-pronominal monosyllables and the Syllable Contact Law

Example	Context
Inne grosse féme éd Tours (Chl'autocar 19) "a fat woman from Tours"	/mdt/
À cho'g grimpètte éd Feuquerolle (Chl'autocar 17) "at the hill of Feuquerolle"	/tdf/
L'oral éd nos éxamins (Chl'autocar 17) "the oral of our exams"	/ldn/
Cho'g granne ligne bleuse éd la Forêt d' Eu (Chl'autocar 34) "the long blue line of the Eu Forest"	/zdl/
J'éroais bél air éd rintrer conme o (Chl'autocar 21) "I would look stupid of coming in like that"	/rdr/
Conme deux gouttes éd ieu (Rinchétte 20) "like two drops of water"	/tdj/

The facts described above provide evidence for the existence of a Clitic Group that is distinct from the Phonological Phrase in Picard. Hayes (1989) proposed this level in his prosodic hierarchy; Nespor & Vogel (1986) provided additional arguments in its favor and proposed the hierarchy in Figure 1.
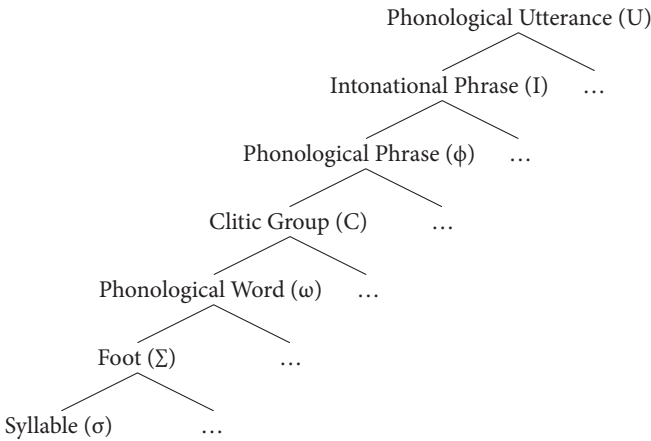
Phonological Utterance (U)

Intonational Phrase (I) …

Phonological Phrase (ϕ) …

Clitic Group (C) …

Phonological Word (ω) …

Foot (Σ) …

Syllable (σ) …

Figure 1. Nespor & Vogel's prosodic hierarchy (source: Horne 1990: 1)

Additional evidence that the clitic group includes only pronominal clitics and negative *n* can be found in the interaction between the bad syllable contact law and vowel epenthesis. As can be seen in (21), bad syllable contacts are tolerated within clitic groups if they involve only two consonants. Specifically, epenthesis is used to avoid a bad syllable contact only if it is required for syllabification purposes.

(21) Bad syllable contact I
    a. *A m' l' o* [am.lɔ] *dit*   (Chl'autocar 82)
        she me it has   said
        "She told me"
    b. *I n' l' avoait* [in.la.vwɛ] *point rconnue*   (Chl'autocar 37)
        he NEG her had   not recognized
        "He had not recognized her"

But what happens if a clitic group that involves a bad syllable contact occurs in a context that creates a need for epenthesis, as illustrated in (22)? Because, in such cases, the problematic sequence of consonants exceeds the domain of the clitic group, epenthesis follows the pattern that applies across word boundaries, i.e., CeCC, and the bad syllable contact is tolerated. This contrasts with the examples in (23), where the bad syllable contact occurs within the clitic group, thus forcing a repair in order to avoid [ʒm] and [tl].

(22) Bad syllable contact II
    a. *ch' est leu qu' éj m' imballe*   [keʒ.mɛ̃.bal]   (Chl'autocar 21)
        it is there COMP I me get-excited
        "that's where I get excited"

b. *pour ét   lancher* [pu.ret.lã.ʃe] *din chés bélles   moaisons*
   for   you sG throw   in the beautiful PL houses
   "in order for you to have access to beautiful houses"   (Chl'autocar 83)

(23) Bad syllable contact III
   a. *J' ém souvarai*   [ʒẽm.su.va.re]   (Chl'autocar 46)
      I me remember.FUT.1SG
      "I will remember"
   b. *ch' est peinne pérdue d' té   l' dire* [tel.dir]   (Chl'autocar 87)
      it is pain   lost   of you sG it say
      "it's not worth telling you"

The necessity of a Clitic Group level that is intermediate between the Phonological Word and the Phonological Phrase has been questioned by a number of researchers, including Horne (1990). Scholars who oppose its addition to the existing prosodic hierarchies argue that it results in an unnecessary proliferation of prosodic levels and that the facts can be captured through analyses that make use of other levels. Vogel (2009) reviews the theoretical and empirical arguments that have been invoked in the literature and concludes that the CG – which she renames the Composite Group, in order to include compounds and, in some languages, certain "level 2" affixes – is required in order to account for phonological processes that are unique to this prosodic level. The facts about vowel epenthesis in Picard provide one more piece of evidence that, in this language at least, we must distinguish the Clitic Group from the Phonological Phrase. The question that remains to be answered is whether the Clitic Group is also different from the Phonological Word.

4.2   Phonological word and clitic group

Picard, like French, allows many complex syllable onsets. Some of these branching onsets are part of the phonological entry of many words (e.g., *trouver* [truve] "find" and *piot* [pjo] "child") and some result from morphological processes. We know that such clusters are branching onsets because, unlike clusters like [km] or [rb], they do not trigger epenthesis when they are preceded by a consonant, as seen in (24a). Similarly, when a verb that starts with a branching onset is preceded by a consonantal prefix or when the future/conditional /r/ suffix is added to a verb stem that ends in /t/, no epenthesis is required, as seen in (24b) and (24c).

(24) Word-internal complex onsets
   a. *point d' trop surprins* [pwẽd.tro.syr.prẽ]   (Chl'autocar 25)
      not   of too surprised
      "not too surprised"

> b. *o   ll'   a   rtrouvè*   [ɔl.lar.tru.vɛ]   (Chl'autocar 23)
>   one him has found-again
>   "he was found"
>
> c. *i   réstroait*   [i.res.trwɛ]   (Chl'autocar 40)
>   he   stay.COND
>   "he would stay"

While morphology can create branching codas, as we just saw, this is not possible inside clitic groups and across word boundaries. (25a) shows that when the *t* clitic precedes the verb *raminteus*, epenthesis obtains. Yet, as (25b) reveals, the sequence [ntr] is pronounceable in Picard. What distinguishes the two examples is that, in *trachoéme*, the complex onset is lexical, whereas in *tu n'té raminteus*, it is not. Similarly, in (25c), the combination of the preposition *d* and the verb *rintrer* triggers epenthesis, even though the same sequence of consonants fails to trigger epenthesis when [dr] is lexical, as in (25d). Thus, clitic groups and phonological phrases both forbid the creation of complex onsets and resort to epenthesis to repair such sequences; however, as can be expected, the epenthesis sites differ: CCeC obtains in the clitic group in (25a), and CeCC obtains at word boundary, as seen in (25c).

> (25) Complex onsets in clitic groups and at word boundaries
>   a. *Tu n'   té   raminteus* [tyn.te.ra.mɛ̃.tø] *point*   (Chl'autocar 107)
>   you NEG you remember   not
>   "You don't remember"
>
>   b. *o   n'   trachoéme* [ɔ̃n.tra.ʃwem]   *méme pu*   (Chl'autocar 25)
>   we NEG sought   even   anymore
>   "we didn't even seek anymore"
>
>   c. *J' éroais   bél   air éd rintrer* [be.lɛ.red.rɛ̃.tre] *conme o*
>   I have.COND beautiful air of enter   like   that
>   "I would look stupid if I came in like that"   (Chl'autocar 20)
>
>   d. *ché sroait   pér driére* [per.dri.er]   (Chl'autocar 79)
>   it   be.COND by   behind
>   "it would be in the backyard"

5.   Discussion and conclusion

This analysis of Picard subject clitics has sought to answer two related but distinct questions: (i) can we explain their (near-)categorical occurrence with lexical and pronominal subjects by the fact that they have been reanalyzed as agreement markers, and (ii) are they still clitics generated in syntax and associated with a verbal host in phonology or have they become lexical affixes?

The reanalysis of subject clitics as agreement markers finds support in the fact that their distribution in contemporary Picard corresponds to that expected of subject agreement markers, as they occur in every construction in which a verb agrees with its subject. This includes the doubling of non-D-linked quantified subjects, their repetition on each tensed verb in VP conjunction, their presence in extraction contexts (subject relative clauses and *wh*-questions), as well as their presence in structures that can be analyzed as cases of stylistic inversion. This analysis is further supported by the fact that the majority of doubled DPs and strong pronouns are shown to occur within the same Intonational Phrase as the verb, as revealed by an analysis of word-initial and word-final epenthesis.

Given that the reanalysis of subject clitics as agreement markers is complete for many speakers of contemporary Picard born after 1930, we also expect them to have been reanalyzed as lexical affixes. However, we have uncovered clear phonological evidence that pronominal clitics differ from lexical affixes by not allowing the creation of complex onsets and that they differ from other types of words, including non-pronominal grammatical monosyllables, by forming prosodic levels which have their own template for vowel epenthesis and in which the constraint against bad syllable contacts is active. Thus, our analysis provides evidence for the existence of the Clitic Group in the prosody hierarchy of Picard; cf. Table 3. While our analysis of the Phonological Word remains tentative, as indicated by the question marks that appear next to *CCeC* and *Active* in that column, the evidence concerning the formation of complex onsets suffices to show that the Phonological Word differs from the Clitic Group.

Table 3. Three distinct prosodic levels in Picard

	Phonological word	Clitic group	Phonological phrase
Epenthesis	CCeC?	CCeC	CeCC
Bad syllable contact	Active?	Active	Inactive
Formation of complex onsets	Allowed	Disallowed	Disallowed

Would an analysis along the same lines provide evidence for clitic groups in colloquial French? Some previous work has concluded that the subject clitics of colloquial French have likewise been reanalyzed as lexical affixes. This conclusion is based, for example, on idiosyncratic forms such as [ʃy] for *je suis* "I am", but not the homophonous form meaning "I follow", and the presence of gaps that are not syntactic in nature (e.g., **Elle me te donne* vs. *Elle me donne à toi* "She gives me to you"; Auger 1994; Culbertson 2010). However, affrication in the Québec variety of colloquial French provides evidence that, in this variety at least, pronominal clitics have not become lexical affixes yet. In Québec French, /t, d/ are realized as

the affricates [tˢ, dᶻ] before high front vowels and glides. Affrication is categorical inside words, as can be seen in (26a)–(26c), but variable across word boundaries, as seen in (26d). The fact that affrication is also variable between a pronominal clitic and its verbal host, cf. (26e) and (26f), suggests that these pronouns are still clitics rather than affixes, as previously claimed.

(26) Affrication in Québec French
    a. *Dîner*   [dᶻine]/*[dine]   "lunch"
    b. *Mature*   [matˢyʁ]/*[matyʁ]   "mature"
    c. *Tuile*   [tˢɥɪl]/*[tɥɪl]   "tile"
    d. *Petite île* [p(ə)tˢɪtˢɪl]/[p(ə)tˢɪtɪl]/*[p(ə)tɪtˢɪl] "small island"
    e. *T'imagines* [tˢimaʒɪn]/[timaʒɪn]   "you.SG imagine"
    f. *Tu t'imagines* [tˢytˢimaʒɪn]/[tˢytimaʒɪn] "you.SG yourself imagine"

The conclusion that subject clitics in Picard are agreement marking clitics challenges two hypotheses concerning the motivation for the grammaticalization of subject clitics into agreement markers. The first hypothesis is that this change is triggered by the loss of agreement-marking suffixes. Indeed, it is well known that, in colloquial French, all three singular persons are homophonous for the majority of tenses and verbs and that, with the replacement of 1PL *nous* by *on*, the only person that is distinct from others is 2PL. This is illustrated with the verb *manger* in the indicative present in Table 4. However, Table 4 also shows that the poor suffixal agreement morphology described for colloquial French does not hold for Picard: whereas all three singular persons are homophonous, the three plural persons are distinct from each other as well as from their singular counterparts. Thus, in Picard, suffixes still play a major role in agreement marking. As a matter of fact, three different persons have identical subject clitics, namely *oz*, and are distinguished through their suffixes: 1PL *o mingeons*, 2PL *o mingez*, and the 3SG indefinite form *o minge* "one eats".

Table 4. "to eat" in the indicative present

Colloquial French	Picard	Gloss
[ʒmɑ̃ʒ]	[ʒmɛ̃ʒ]	"I eat"
[tymɑ̃ʒ]	[tymɛ̃ʒ]	"you.SG eat"
[imɑ̃ʒ]	[imɛ̃ʒ]	"he eats"
[ɔ̃mɑ̃ʒ]	[õmɛ̃ʒ]	"one eats"
[ɔ̃mɑ̃ʒ]	[õmɛ̃ʒɔ̃]	"we eat"
[vumɑ̃ʒe]	[õmɛ̃ʒe]	"you.PL eat"
[imɑ̃ʒ]	[imɛ̃ʒt]	"they eat"

The second hypothesis is that "the transformation of subject-clitic pronouns into affixes and the rise of *ne* deletion were parallel developments and that such a transformation has played a role in the rise of *ne* deletion" (Martineau & Mougeon 2003: 143), a hypothesis that is supported by the fact that Canadian French has near categorical deletion of *ne* and that some of its varieties have high rates of subject doubling, while the Walloon dialects described by Remacle (1960) maintain *ne* and lack subject doubling (Martineau & Mougeon 2003: 144). Once again, Picard provides evidence that such a relation need not hold, as Auger & Villeneuve (2008: 241) have shown that in the Vimeu variety of Picard, *ne* deletion is infrequent in writing, with an average of 4.2% for six different authors and individual frequencies ranging from 0.3% to 7.2%,[6] thereby illustrating the importance of taking into account Gallo-Romance varieties such as Picard as tests of the evolutionary cycle involving subject clitic pronouns and *ne* deletion hypothesized by Martineau & Mougeon (2003).

The analysis of subject clitics in the Vimeu variety of Picard, which was initially motivated by our attempt to determine what French might look like in the absence of strong normative pressure from teachers and other prescriptive figures, has provided clear evidence that these elements have been reanalyzed as agreement markers. However, contrary to our expectation, these elements still behave differently from lexical affixes. Furthermore, they also behave differently from other grammatical monosyllables. Our analysis of vowel epenthesis at different levels of prosodic structure and the formation of branching codas has led us to conclude that this language provides strong evidence for the existence of the Clitic Group. This type of analysis is only one example of what research on endangered regional languages and dialects can teach us about variation and change within the Gallo-Romance family and of the great urgency of conducting more such research before such varieties disappear.

---

6. The higher *ne* deletion rate of 49% reported for oral Picard in Villeneuve & Auger (2013: 120) is interpreted as evidence of interference of colloquial French into oral Picard.

References

Ashby, William J. 1977. *Clitic inflection in French; An historical perspective*. Amsterdam: Rodopi.

Ashby, William J. 1980. Prefixed conjugation in Parisian French. In Herbert J. Izzo (ed.), *Italic and Romance linguistic studies in honor of Ernst Pulgram*, 195–207. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.18.24ash

Ashby, William J. 1988. The syntax, pragmatics, and sociolinguistics of left- and right-dislocations in French. *Lingua* 75. 203–229. https://doi.org/10.1016/0024-3841(88)90032-0

Auger, Julie. 1994. Pronominal clitics in Colloquial Québec French: A morphological approach. Philadelphia, PA: University of Pennsylvania dissertation.

Auger, Julie. 2000. Phonology, variation, and prosodic structure: Word-final epenthesis in Vimeu Picard. In Josep M. Fontana, Louise McNally, M. Teresa Turell & Enric Vallduvi (eds.), *Proceedings of the First International Conference on Language Variation in Europe (ICLaVE)*, 14–24. Barcelona: Universitat Pompeu Fabra.

Auger, Julie. 2001. Phonological variation and Optimality Theory: Evidence from word-initial vowel epenthesis in Picard. *Language Variation and Change* 13(3). 253–303. https://doi.org/10.1017/S0954394501133016

Auger, Julie. 2003a. Le redoublement des sujets en picard. *Journal of French Language Studies* 13(3). 381–404. https://doi.org/10.1017/S0959269503001200

Auger, Julie. 2003b. The development of a literary standard: The case of Picard in Vimeu-Ponthieu, France. In Brian D. Joseph, Johanna DeStefano, Neil G. Jacobs & Isle Lehiste (eds.), *When languages collide: Perspectives on language conflict, language competition, and language coexistence*, 141–164. Columbus, OH: Ohio State University Press.

Auger, Julie. 2003c. Les pronoms clitiques sujets en picard: une analyse au confluent de la phonologie, de la morphologie et de la syntaxe. *Journal of French Language Studies* 13(1). 1–22. https://doi.org/10.1017/S0959269503001066

Auger, Julie. 2018. Revitalisation, variation régionale et purisme dans le développement d'un standard littéraire: le cas du picard. In Christine Meyer & Paula Prescod (eds.), *Langues choisies, langues sauvées: poétiques de la résistance*, 155–171. Würzburg, Germany: Königshausen & Neumann.

Auger, Julie & Ryan Hendrickson. 2011. Picard verbal morphology: What it tells us about syllable structure. Paper presented at the CUNY Conference on the Phonology of Endangered Languages, New York City, NY, January 12–14.

Auger, Julie & Anne-José Villeneuve. 2008. *Ne* deletion in Picard and in regional French: Evidence for distinct grammars. In Miriam Meyerhoff & Naomi Nagy (eds.), *Social lives in language – Sociolinguistics and multilingual speech communities*, 223–247. Amsterdam: John Benjamins. https://doi.org/10.1075/impact.24.16aug

Auger, Julie & Anne-José Villeneuve. 2010. La double expression des sujets en français saguenéen: étude variationniste. In Wim Remysen & Diane Vincent (eds.), *Hétérogénéité et homogénéité dans les pratiques langagières: mélanges offerts à Denise Deshaies*, 67–86. Québec: Presses de l'Université Laval.

Barnes, Betsy. 1985. *The pragmatics of left detachment in spoken standard French*. Amsterdam: John Benjamins. https://doi.org/10.1075/pb.vi.3

Beaulieu, Louise & Patricia Balcom. 1998. Le statut des pronoms personnels sujets en français acadien du nord-est du Nouveau-Brunswick. *Linguistica atlantica* 20. 1–27.

Brandi, Luciana & Patrizia Cordin. 1989. Two Italian dialects and the null subject parameter. In Osvaldo Jaeggli & Ken Safir (eds.), *The null subject parameter and parametric theory*, 111–142. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-009-2540-3_4

Burnett, Heather & Julie Auger. 2016. Syntax, semantics and affect in Picard secondary negation. Paper presented at Linguistic Symposium on the Romance Languages (LSRL) 46, Stony Brook University, Stony Brook, NY, March 31–April 3.

Campion, Elizabeth. 1984. Left dislocation in Montréal French. Philadelphia, PA: University of Pennsylvania dissertation.

Clements, George N. 1990. The role of the sonority cycle in core syllabification. In John Kingston & Mary E. Beckman (eds.), *Papers in laboratory phonology*, 283–333. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511627736.017

Cochet, E. 1933. *Le patois de Gondecourt (Nord)*. Paris: Droz.

Coveney, Aidan. 2003. Le redoublement du sujet en français parlé: une approche variationniste. In Anita Berit Hansen & Maj-Britt Mosegaard Hansen (eds.), *Structures linguistiques et interactionnelles dans le français parlé: actes du colloque international, Université de Copenhague du 22 au 23 juin 2001*, 111–143. Copenhagen: Museum Tusculanum Press.

Coveney, Aidan. 2005. Subject doubling in spoken French: A sociolinguistic approach. *The French Review* 79(1). 96–111.

Culbertson, Jennifer. 2010. Convergent evidence for categorial change in French: From subject clitic to agreement marker. *Language* 86(1). 85–132. https://doi.org/10.1353/lan.0.0183

Dagnac, Anne. 2015. 'Pas', 'mie', 'point' et autres riens : de la négation verbale en picard. In Jan Goes & Mariana Pitar (eds.), *La négation en français : études linguistiques, pragmatiques et didactiques*, 129–152. Artois: Artois Presse Université. https://doi.org/10.4000/books.apu.6786

Dauby, Jean. 1979. *Le livre du 'rouchi' parler picard de Valenciennes*. Amiens: Musée de Picardie.

Dawson, Alain. 2008. "Bienvenue chez les Chtis": la langue opaque. *Langues et cité* 12(4).

Dawson, Alain. 2010. *Guide de conversation picard de poche*. Paris: Asimil.

De Cat, Cécile. 2007. *French Dislocation*. Oxford: Oxford University Press.

Debrie, René. 1974. Étude linguistique du patois de l'Amiénois. Amiens: Archives départementales de la Somme. Paris : Université de Paris dissertation (1960).

Edmont, Edmond. 1897. *Lexique Saint-Polois*. Geneva: Slatkine Reprints. (Reprint of the original 1897 self-published edition).

Fonseca-Greber, Bonnibeth Beale. 2000. The change from pronoun to clitic to prefix and the rise of null subjects in spoken Swiss French. Tucson, AZ: University of Arizona dissertation.

Fuβ, Eric. 2005. *Rise of agreement. A formal approach to the syntax and grammaticalization of verbal inflection*. Amsterdam: John Benjamins. https://doi.org/10.1075/la.81

Gaglia, Sascha & Christoph Schwarze. 2015. The controversial status of Romance pronominal clitics – a new criterion. *Linguistische Berichte* 242. 103–139.

Hayes, Bruce. 1989. The prosodic hierarchy in meter. In Paul Kiparsky & Gilbert Youmans (eds.), *Phonetics and Phonology; Rhythm and Meter*, Volume 1, 201–260. Cambridge, MA: Academic Press. https://doi.org/10.1016/B978-0-12-409340-9.50013-9

Heap, David, Michèle Oliviéri & Katerina Palasis. 2017. Clitic pronouns. In Andreas Dufter & Elisabeth Stark (eds.), *Manual of Romance morphosyntax and syntax*, 183–229. Berlin: De Gruyter. https://doi.org/10.1515/9783110377088-005

Horne, Merle. 1990. The clitic group as a prosodic category in old French. *Lingua* (82)1. 1–13. https://doi.org/10.1016/0024-3841(90)90053-N

Hrkal, Ed. 1910. Grammaire historique du patois picard de Démuin. *Revue de philologie française et de littérature* 24. 118–140, 175–204, 241–277.

Kayne, Richard & Jean-Yves Pollock. 1978. Stylistic inversion, successive cyclicity, and Move NP in French. *Linguistic Inquiry* 9(4). 595–621.

King, Ruth & Terry Nadasdi. 1997. Left dislocation, number marking, and (non-)standard French. *Probus* 9. 267–284. https://doi.org/10.1515/prbs.1997.9.3.267

Landrecies, Jacques. 2006. Astérix, Cafougnette, Tintin et Martine ou le picard en tête de gondole. *Nord'* 48. 63–90.

Ledieu, Alcius. 1909. *Petite grammaire du patois picard*. Paris: H. Welter.

Legendre, Géraldine, Jenny Culbertson, Isabelle Barrière, Thierry Nazzi, & Louise Goyet. 2010. Experimental and empirical evidence for the status and acquisition of subject clitics and agreement marking in adult and child spoken French. In Vincent Torrens, Linda Escobar, Anna Gavarró, & Juncal Gutiérrez (eds.), *Movement and clitics*, 333–360. Newcastle: Cambridge Scholars Publishing.

Martineau, France & Raymond Mougeon. 2003. A sociolinguistic study of the origins of *ne* deletion in European and Quebec French. *Language* 79(1). 118–152. https://doi.org/10.1353/lan.2003.0090

Nadasdi, Terry. 1995. Subject NP doubling, matching and minority French. *Language Variation and Change* 7(1). 1–14. https://doi.org/10.1017/S0954394500000879

Nadasdi, Terry. 2000. *Variation grammaticale et langue minoritaire: le cas des pronoms clitiques en français ontarien*. Munich: LINCOM Europa.

Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris.

Oudin, Antoine. 1632/1970. *Grammaire française rapportée du langage du temps*. Geneva: Slatkine Reprints.

Palasis, Katerina. 2013. The case for diglossia: Describing the emergence of two grammars in the early acquisition of metropolitan French. *Journal of French Language Studies* 23(1). 17–35. https://doi.org/10.1017/S0959269512000348

Palasis, Katerina. 2015. Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua* 161. 125–143. https://doi.org/10.1016/j.lingua.2014.11.012

Poletto, Cecilia. 2000. *The higher functional field: Evidence from Northern Italian dialects*. Oxford: Oxford University Press.

Remacle, Louis. 1960. *Syntaxe du parler wallon de la Gleize*. 3 volumes. Paris: Belles Lettres.

Rizzi, Luigi. 1986. On the status of subject clitics in Romance. In Osvaldo Jaeggli & Carmen Silva-Corvalan (eds.), *Studies in Romance Linguistics*, 391–419. Dordrecht: Foris. https://doi.org/10.1515/9783110878516-025

Roberge, Yves. 1990. *The syntactic recoverability of null arguments*. Montréal: McGill-Queen's University Press.

Sankoff, Gillian. 1982. Usage linguistique et grammaticalisation: les clitiques sujets en français. In Norbert Dittmar & Brigitte Schlieben-Lange (eds.), *Sociolinguistique dans les pays de langue romane*, 81–85. Tübingen: Gunter Narr.

Suñer, Margarita. 2003. The lexical preverbal subject in a Romance null subject language. Where are thou?. In Rafael Núñez-Cedeño, Luis López & Richard Cameron (eds.), *Romance perspective on language knowledge and use: Selected papers from the 31st Linguistic Symposium on Romance Languages (LSRL), Chicago, 19–22 April 2001*, 341–357. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.238.25sun

Vasseur, Gaston. 1996. *Grammaire des parlers picards du Vimeu (Somme)*. Abbeville: F. Paillart.

Vaugelas, Claude Favre de. 1647. *Remarques sur la langue françoise utiles à ceus quiveulent bien parler et bien escrire*. Paris: J. Camusat & P. le Petit (facsimile edition by Jeanne Streicher, Paris: Slatkine Reprints, 1934)

Villeneuve, Anne-José & Julie Auger. 2013. *Chtileu qu'i m'freumereu m'bouque i n'est point coér au monne:* Grammatical variation and diglossia in Picardie. *Journal of French Language Studies* 23(1). 109–133. https://doi.org/10.1017/S0959269512000385

Vogel, Irene. 2009. The status of the clitic group. In Janet Grijzenhout (ed.), *Interface explorations: Phonological domains: Universals and deviations*, 15–46. Berlin: Mouton de Gruyer. https://doi.org/10.1515/9783110219234.1.15

Zribi-Hertz, Anne. 1994. La syntaxe des clitiques nominatifs en français standard et en français avancé. *Travaux de linguistique et de philologie* XXXII. 131–147.

Data sources

[Chl'autocar] Leclercq, Jean. 1996. *Chl'autocar du Bourq-éd-Eut*. Abbeville: Ch'Lanchron.

[Deglicourt] Deglicourt, Pierre. 1987. Un 31 d'octobe coeud…. *Ch'Lanchron* 29. 22–23.

[Dufrêne] Dufrêne, Jules. 1991. Quante o n'o pu qu'ses zius… pour braire. *Ch'Lanchron* 43. 10.

[JVasseur] Vasseur, Jehan. 1996. I n'feut point gadrouilleu ch'pain…. *Ch'Lanchron* 63. 8.

[Lettes] Vasseur, Gaston. 2003. *Lettes à min cousin Polyte*. Abbeville: F. Paillart.

[Réderies] Lecat, Charles. 1977. *Réderies*. Fressenneville: Imprimerie Carré.

[Rinchétte] Chivot, Eugène. 1993. *Rinchétte*. Abbeville: Ch'Lanchron

[Viu temps] Vasseur, Gaston. 1969. *Histoéres du viu temps*. Abbeville: Imprimerie Lafosse.

A child's view of Romance modification

Ana T. Pérez-Leroux
University of Toronto

Languages vary as to what kind of phrasal categories allow recursive iteration of self-same embedding. Children first learn an embedding rule, then must learn whether the rule can apply recursively or not. However, direct experience of recursive embedding is rare in the input. A study of recursive nominal modification in Spanish show children acquire different types of modification (possession, part-whole relations) at different times even if these are expressed with the same preposition *de*. This suggest that the domain of rule formulation is narrower than syntactic category (PPs) or even lexical particle (*de*). Bilingual children show delays in acquiring a first level of embedding rule but not in allowing the rule to be recursive. This suggests that learning recursive modification is not sensitive to the concomitant reductions in input in bilingual contexts. I argue that children learn that embedding rules are recursive by inference from the productivity of simple embedding rules. The evidence on the acquisition of recursive nominal modification points to the limitations of the parameter setting model of syntactic development.

Keywords: modification, recursion, syntactic development, Spanish, productivity, DP

1. Introduction

The last decades of research in generative syntax have uncovered much about the structure of the noun phrase in Romance languages, including issues of definiteness, adjectives, nominalizations, and noun ellipsis, among others. Some mysteries remain. Consider the unexplained restrictions on the distribution of prepositional modifiers in Spanish: the set of lexical prepositions that embed prepositional phrase (PP) as noun modifiers is quite limited (compared to the range of PP modifiers in the verbal domain), following some ill-defined semantic constraints. With these constraints as a backdrop, I present new data on the acquisition of recursive PP modification in Spanish, and on children's production of PP modification and

recursion. This data serves to frame a comparison between two different notions of evidence: the kind of evidence that settles theoretical debates and plausible learnability evidence.

2.   UG then and now: How generative approaches neglected grammatical learning

By conceptualizing cognition as symbol manipulation, the cognitive revolution put language at the center of its enterprise. Chomsky held that language should be studied from the perspective of the unbounded capacity speakers have for comprehending and producing an infinite number of utterances rather from the perspective of its extensions, the corpus that results from the externalization of this capacity (Chomsky 1957, 1959). The recursive application of embedding rules illustrates the unboundedness of language.

(1)   The dog that chased the cat that spied on the bird that …

What children learn then is neither a verbal behavior nor a rich set of expressions, but a system of rules (the example in (1) is built on the basis of a relativization rule). Language acquisition was thus initially conceived as a process of rule discovery and formulation (Berko 1958; Brown & Berko 1960). This perspective was superseded by a richly articulated view of the innate disposition to learn languages, the theory of Universal Grammar (UG) (Chomsky 1981, 1986). Under a UG approach, a child has innate knowledge of universal principles of language structure, i.e., the commonalities across languages, and of parameters, the possible patterns of language variation. By detecting certain key properties of utterances linked to a parameter (i.e., the triggering experience) learners presumably accessed a host of associated (but less ostensive) grammatical properties. Learning (in the sense of acquiring language-specific information rather than accessing universal representations) was construed in terms of learning words and morphemes, and parameter-setting, which represented discontinuous stages in grammatical development except under grammar competition models, such as Yang (2004).

Over the years, as the study of cognition shifted focus from symbolic systems to statistical methods, the centrality of linguistic theory in the cognitive enterprise diminished. Field-internal challenges also arose to formal (symbolic) universals, including the claim that recursive embedding is a universal property (Evans & Levinson 2009). Within language acquisition, emergentist critics of the generative approach focus on the interaction between experience and domain-general learning abilities (Pine & Lieven 1997; Tomasello 2000), claiming that learning is usage-based and that parameter-setting approaches cannot model gradual learning.

Claims about the lexical nature of early language has not stood statistical analyses (Yang 2004, 2016; Valian, Solt & Stewart 2009), so the question of gradual development deserves more discussion.

While usage-based approaches are considered insufficient within the developmental field (Ninio 2011), the case for parameter setting as a learning model is comparably weak. The evidence is clear that children are sensitive to principles of structural organization (Lidz, Waxman & Freedman 2003; Syrett & Lidz 2009; Roeper & de Villiers 2011; Hunsicker & Goldin-Meadow 2012, *inter alia*). Despite the lack of concrete empirical evidence in support of parameter setting, parameters continue to be generally accepted (Pearl & Lidz 2013). David Lightfoot was early to point to a general failure to identify the triggering experience: " … surprising that little attention has been paid to what it takes to set these parameters. Sometimes this lack of attention undermines the claims being made" (1991: ix). No evidence was found for morphology as trigger. For n-drop and gender, corpus data suggests at most that morphological mastery precedes syntax; not that it 'triggers' it (Snyder, Senghas & Ihnman 2001). Parametrically-related properties occasionally emerge together in spontaneous speech (Snyder 2007); but such temporal clusterings are not the norm. At their best, parameters describe contingencies between features not predictable on the basis of surface characteristics, as in Han, Lidz & Musolino (2016) where it is shown that population-level variation is not based on individual variability. Children make categorical parametric decisions about the scope of negation in Korean that are unrelated to parental patterns. Yet, no trigger is proposed, so that parameters remain as useful characterization of static properties of grammars and grammatical stages, but do not advance our understanding of the dynamics of developmental changes. Parameters are best at describing the outcome of learning rather than the process. For this, we must go elsewhere.

Theoretical linguistics has been generally slow to integrate linguistic constraints with statistical learning. Some learnability models give a role to statistical learning when implementing parametric learning, as grammar selection via parsing (Fodor 1998; Yang 2002). More can be said, given what is known about how infants can extract generalizations from the statistical frequencies in the input (Saffran, Aslin & Newport 1996; Marcus et al. 1999). Statistical learning is a domain-general ability that supports a wide range of acquisition phenomena, including learning word boundaries, and phonetic and grammatical categories. Distributional learning transforms the input into useable representations, but distributional information is not a grammar. Many constraints on grammatical systems cannot be extracted from surface frequencies. Therefore, statistical learning alone cannot provide a full alternative to nativist explanations for linguistic constraints. Learners need to map the output of statistical learning into symbolic representations (Marcus et al. 1999), which obey general (universal) principles of structural organization.

The principal challenge for translating statistical learning into grammar is how to associate possible forms and possible meanings. Any explanatory theory needs mechanisms for linking distributional analyses and potential interpretations (the mapping problem); and for reining in generalizations about the various distributions (the constraints problem).

This requires a narrower approach to learning than traditional parameter-setting approaches can accommodate. To start with the mapping problem, the obvious solution is to apply the syntactic bootstrapping mechanism, initially proposed for verb learning (Gleitman 1990; Gleitman et al. 2005), to the elements in the functional inventory. Syntactic Bootstrapping is a probabilistic multiple-cue learning process which extracts semantic information from a distributional environment of lexical items (i.e., subcategorization contexts), building up increasingly detailed, highly structured lexical representations, by mapping sentence structure to world situations. Despite its usefulness, there is little explicit discussion about how to apply it to grammar (for some exceptions, see Valian et al. 2009 and Sneed 2008).

Reining in generalizations is a different problem. Appeals to underlying universal principles is inappropriate, since most of the relevant constraints are not universal. Consider the acquisition of determiner meanings, which depend not just on the other determiners, but also on the availability of null forms (bare nouns). Chierchia's (1998) parameter aimed to explain the semantic distribution of bare nouns and DPs crosslinguistically (Guasti et al. 2008). This parameter relied on the insight that the function or meaning of forms depends on the other forms in the language (de Saussure 1916). This assumption is widespread in acquisition: we see it invoked for the lexicon in the Principle of Contrast (Clark 1995) and the Mutual Exclusivity Constraint (Markman & Wachtel 1988), for morphosyntax in the Morphological Uniqueness Principle (Pinker 1984) and Pre-emption (Brooks & Tomasello 1999), and in formal semantics in Maximizing Presupposition (Heim 1991; Sauerland 2006). Can parameters achieve these same effects?

Yang (2016) offers the first computationally explicit modelling of form competition that goes beyond surface distributions. The frequency distributions of words and word combinations (n-grams, determiner+noun sequences, etc.) follow a Zipfian curve, where the frequency of an element is the inverse of its rank frequency order. Considering the trade-off between space and time complexity in accessing a form, i.e., whether storage is lexically listed, or rule-based, Yang proposes the Tolerance Principle, an equation predicting how many exceptions are possible before a grammatical rule is considered not productive. The Tolerance Principle is also a hypothesis about how children map impoverished input to a rule system. The principle is broad: any detectable pattern (distributional, phonological, or semantic) will be formulated as a rule. It is also recursive: the productivity of a rule can be reassessed with additional input. Limiting the scope of generalizations

this way eliminates the need for negative evidence. Learners can incorporate constraints on generalizations below category level, such as adjectival subclasses (Yang 2015). Crucially, the process operates solely from positive evidence, reducing the need to attribute negative knowledge to universal constraints. Distributional input maps onto symbolic, contrastive rules, with economy principles constraining the powerful rule-making capacities. This allows learning language-specific rules (either broad or lexically narrow) without invoking parameters. The semantic aspects of rule learning would presumably be bootstrapped from available sentence-level information.

3.   Recursion as a learning problem

Recursive embedding requires generalization from poor stimuli. Iterated embedding is not frequent in spontaneous speech to children (Roeper & Snyder 2005). Interest in the acquisition surged after Hauser, Chomsky & Fitch's (2002) claim that recursion is the unique property of human languages (Hauser et al. 2002; Sauerland & Gärtner 2007), and functionalists' objections (Evans & Levinson 2009), on the basis of the Pirahã exceptionality claim (Everett's 2005; see responses in Pesetsky 2009 and Nevins, Pesetsky & Rodriguez 2009). The debates center on the acknowledged existence of variation in phrasal embedding across languages. The focus is on recursive structures (i.e., the unbounded iteration of self-embedding of categories of the same type, and with identical categorial labels and distributional properties), which rely on but differ from the formal property of recursion (Widmer et al. 2017). Given variation, recursive structures need to be learned: either as a 'parameter' or as a 'rule', arising from distributional learning and constrained by it.

   Roeper & Snyder (2005) point out Saxon genitive possessors exist in German and English as in (2), but are only recursive in the latter. The English learner must learn that the possessive rule can be recursive as in (3) whereas the German child must avoid doing so (Pérez-Leroux et al. forthcoming). Noun-noun compounds is another case of variation. They appear in English and Spanish but are only recursive in English as in (4a). A child might learn two N-N compounds in Spanish but must not generalize them to a recursive structure as in (4b–c).

(2)   *Marias Haus*
      Maria's house

(3)   Maria's cousin's neighbour's house, etc.

(4)   a.   Christmas tree/Christmas tree cookie…
      b.   *Mujer araña   Batman mobíl*
           woman-spider batman-mobile

    c. [**]Mujer araña mobíl
        woman spider mobile

Roeper (2011) suggests that children first learn that English compounds are productive, then learn recursivity as a separate step. If recursive structures are (generally) universal, the ability to learn them should be robust. Goldin-Meadow (1982) points that some properties can withstand variation in the learning conditions, whereas others require robust experience. She argued that recursive embedding and hierarchical structure are resilient properties of language, showing that recursive, endocentric NPs emerge spontaneously in the home signing systems created by deaf children (Hunsicker & Goldin-Meadow 2012).

    Diachronic data supports this notion. The Indo-European family exhibits substantive synchronic and diachronic diversity in the types of recursive NP structures, such as genitives, adpositions, case, and others. Widmer et al. (2017) noted that the types that allow recursion emerge and disappear in relatively short time spans (in diachronic terms). When one recursive form disappears in an Indo-European language, another increases in use. These results suggest recursive phrasal embedding is resilient, not an accidental property of some languages. According to Fitch (2010), economy principles are involved. Processing economy favors recursive rules, as a single rule is specified for all levels of embedding. However, acknowledging a bias towards learning phrasal recursion does not answer how learners decide which forms are possible. Do children need actual examples to learn a type of phrase is recursively embedded, and if so, how many? Are other features of the input relevant? The next section discusses the learnability conditions for recursive NP modification in Spanish, and present new data from collaborative work.

4.   Acquiring recursive NP modification in Spanish

4.1   Recursion in child language

Children do not spontaneously use multiple modification until well into their school years (Eisenberg et al. 2008). In experiments, Kindergarten-aged children show comprehension difficulties; they assign coordinate interpretations to recursive structures (Roeper & Snyder 2005; Limbach & Adone 2010; Roeper 2011, etc.). Elicited production data indicate that embedding is the locus of difficulty: preschoolers easily produce three coordinated NPs (Pérez-Leroux et al. 2012), but struggle with three NPs when it is a noun with recursive modifiers. Both children and adults find recursive sequences significantly more challenging than non-recursive sequences (Antonsen & Farhang 2018; Pérez-Leroux & Roberge 2018).

Can input explain these delays? Possibly. Roeper and Snyder (2005) report that recursive expressions are scarce in parental speech and non-existent in children's, but do not report actual counts. Data by Pettibone (in preparation) find limited input on NP modification in longitudinal corpora in Spanish (Table 1). Simple NP internal modification is generally infrequent. Instances of double modification, only a subset of which are recursive, are rarer still by an order of magnitude. For comparison, I include the Pirahã numbers in Futrell et al. (2016). The frequency of recursive embedding in Spanish parental data is not significantly different from the zero frequency given for Pirahã ($\chi^2 = .088$, $p = .76$ for Juan's data; $\chi^2 < .001$, $p = 1$ for María's). If low prevalence can be considered as evidence of absence in field linguistics, it can certainly be admitted as evidence of poverty of stimulus in developmental linguistics.

Table 1. Frequency of NP modification in parental input in Spanish, compared to reports on Pirahã

Modification type	Juan's input 21,654 words	Maria's input 86,436 words	Pirahã 6830 words
Adjectives	79	89	
PP (all lexical Ps)	133	82	(29 possessors)
RC	24	39	(2 ambiguous)
Recursive	3	3	0 (1 ambiguous)
Double non recursive	7	4	

If children are waiting for good input of phrasal recursion, they might wait so long that it disappears at the level of the language. The Indo-European case study shows that recursive NPs stay, but their forms change. Children are thus likely to recruit information from available forms. Considering NP structure from the start of what acquisition researchers call "first syntax", I propose three stages: (1) Simple DP acquisition, (2) Acquisition of (Level 1) embedding rules, (3) Acquisition of phrasal recursion (Level 2+, or rule iteration).

At the first stage, children map the distribution of adnominal elements, unravelling the functional layers of DPs (Panneman 2007). This happens very early in Spanish (Pérez-Leroux & Castilla-Earls 2016). The second stage involves NP elaboration, when children learn the morphemes that link nouns and their modifiers (including case, prepositions, and relative pronouns). Rule acquisition involves knowledge of the embedding and embedded categories, and the linker that connects them, as well as any relevant distributional constraints between the three elements. We see the third stage not as a distinct step, but as a generalization of the Stage 2 rules. If instances of recursive embedding are too rare to be reliable, learnability-wise, we must consider what other properties of the ambient language

can support the inference that a rule can be recursive. As mentioned above, Spanish N-N compounds exist but are not productive; German possessive *-s* is not recursive, but neither it is productive. Productivity offers a natural limit for phrasal recursion. A child will treat rule X as recursive, provided sufficient evidence that X is productive (in Yang's 2016 sense) and that the child has the capacity to process/produce such complex structures. No role is given to direct experience with actual recursive phrases. There is no need to posit a stage where doubly embedded sequences will be lexically stored, as usage-based approaches might predict.

4.2   NP internal modifiers in Spanish

Spanish is a good place to examine children's acquisition of recursive modification, because it has a relatively simple inventory of modifiers: relative clauses (RCs), predicative structures formed using the relator/linker particle *de*, and PPs that use restricted lexical prepositions. Although grammarians seldom differentiate between *de* and other prepositions, their distributions are vastly different. Like RCs, *de*-phrases are essentially unrestricted, whereas other Spanish PP modifiers are subject to various limitations. These constraints have not been studied beyond some basic descriptive statements:

i.   Lexical prepositions vary as to where they can be used in modifier contexts. (Moreira-Rodriguez 2006; Picallo 2012)
ii. Locative prepositions are generally excluded. They occur in varieties of US Spanish, but are excluded in monolingual contexts. They can appear in topic-comment structures such as titles, captions and exclamations. (Moreira-Rodriguez 2006) (6)
iii. Deverbal Ns can inherit PP arguments (Picallo 2012) (Example (7))
iv. Alienable possession is blocked for comitative prepositions (8)
v.   Stage-level predicates cannot function as PP modifiers (9)

(5) *De>con/sin> para >a*   (Moreira-Rodriguez 2006)

(6) a.   *El puente sobre el río Kwai*
          "Bridge over the river Kwai"
     b.   **El botón en su pantalla*   (M. C. Cuervo, p.c.)
          "the button on your screen"
     c.   *¿La mujer al lado de mi padre? Su secretaria.*
          "The woman next to my father? His secretary."
     d.   *¡Una serpiente en el cuarto de baño!*
          "A snake in the bathroom!"   (Moreira-Rodriguez 2006)

(7) a. *El camino hacia Toledo*
"The road [*lit*. walk] to Toledo"
b. *El discurso sobre la poesía de Lorca*
"The talk about Lorca's poetry"

(8) *La mujer con lentes* / *La mesa con vasos
"The woman with glasses" / "The table with glasses"

(9) a. *El puente sobre el rio Kwai/*
"Bridge over the river Kwai"
a′. *el lápiz sobre el libro
"the pencil over the book"
b. *La pared entre los dos cuartos es muy fina*
"The wall between the two rooms is very narrow."
b′. *La niña entre los dos árboles es mi prima
"The little girl between the two trees is my cousin"

When a given modifier is not acceptable, the same sense can be expressed via a relative clause (c.f., 10a–10b); alternatively, a vague/unspecified equivalent can be formed with *de* (10c):

(10) a. *El lápiz sobre el libro.
b. *El lápiz que está sobre el libro*
c. *El lápiz del libro*
"The pencil on the book"

Spanish children must learn that relative clauses and *de*-phrases are unrestricted and can be recursive; we predicted they should not extend nominal recursion to other prepositions. However, if a lexical preposition is productive at Level 1 for a speaker, it should also be recursive. This possibility that can be explored in the context of the above hierarchy.

4.3   The complexity and recursion project

The complexity and recursion project (Pérez-Leroux & Roberge 2014) aimed to examine the acquisition of NP recursion in five languages: Japanese, German, English, French and Spanish, using a quasi-experimental referential elicitation tool initially developed in Pérez-Leroux et al. (2012), and extended to other structures (in collaboration with A. Castilla-Earls, S. Béjar and D. Massam). A set of images presented double sets of contrasting entities. In the scene below we see boxes, cans and tomatoes. The target referent is best described with two instances of modification (*box of cans/cans of tomatoes*).

Figure 1. Elicitation trial for relational nouns

(11)   Prompt: *¿Dónde está el ratón?*
         "Where is the mouse?"
       Target: *Sentado encima de <u>la caja de latas de tomates</u>*
         "Sitting on top of the box of the cans of tomatoes."

The scene in (11) elicited relational noun modifiers; other types of modifiers were also targeted: possessives, comitatives (accompaniment), and locatives. There were 6 trials per condition, interspersed with a variety of distractors and with one double non-recursive elicitation condition.

(12)   *La cometa del hijo del bombero.*
        "The fireman's boy's kite"   (possessive)

(13)   *El bebé (que está) con la mujer (que está) con unas flores.*
        "The baby with the woman with the flowers"   (comitative)

(14)   *El pájaro (que está) encima del cocodrilo (que está) en el agua*
        "The bird (that is) on the crocodile (that is) in the water"   (locative)

4.4   The Colombia study

Spanish data collection was directed by A. Castilla-Earls and M. F. Lara Díaz. Participants were 112 children and 22 adult controls from Bogotá, Colombia. Two children were identified as atypical and not included in the report. Children were classified by age: four-year-olds (mean age 4;06, *n* = 28); five-year-olds (mean age 5;05) and six-year-olds (mean age 6;03, *n* = 30). Children completed the recursion tasks, plus additional developmental tests of vocabulary, general language, memory and intelligence, reported in detail in Pérez-Leroux et al. (in preparation). We explored the following questions:

–   What is the developmental timeline of recursive nominal modification in Spanish? To what extent are child patterns comparable to those of adults?
–   Do children and adults have comparable rates of success across the various conditions?
–   Are children equally likely to use the different forms of recursive embedding as adults?

Target responses had to (a) successfully describe the referent and (b) contain recursive use any of the target embedding strategies (*de* (15a), RCs (15b) or their combinations (15c)). Non-target responses included incomplete descriptions such as simple NPs and phrases with a single modifier (15d), or described the referent by alternative means as in (15e), which, while correct, do not document the acquisition of recursive structures.

(15) a. *la del amigo de Dora* (4EB 4;09)
"The (one) of-the friend of Dora"

b. *la niña que tiene el perro que tiene sombrero* (6JB 6;01)
"The girl that has the dog that has hat"

c. *la que tiene el perro con sombrero* (5AP 5;07)
"The (one) that has the dog with hat"

d. *la cometa…del niño* (5MR; 5;00)
"the kite…of-the kid"

e. *el de la derecha* (5VM; 5;07)
"the (one) on the right"

Figure 2 shows mean target responses per condition for groups. Children produce significantly less target recursive NPs than adults. Four- and five-year-olds have low rates of targets, but these increase significantly by age six. Speakers gave more recursive targets in response to possessives and comitative contexts, less for
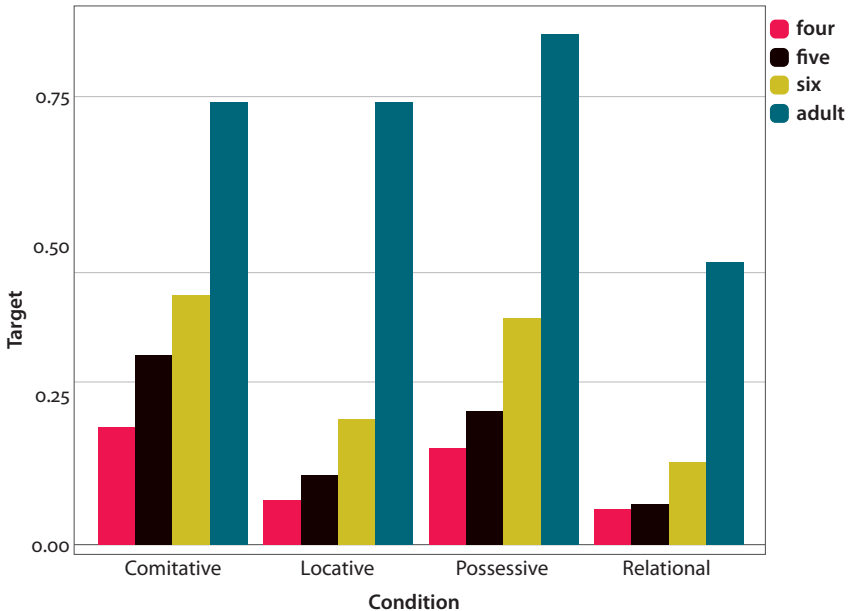

Figure 2. Proportion of targets per condition for the four age groups in the Colombia study

locatives and even less for relational nouns. There was little interaction between age and condition.

Table 2 shows the most common linkers used in simple (Level 1) descriptions, which are RCs closely followed by *de*. We also observed few comitatives and spatial prepositions. There are important differences with the distribution of linkers used in target responses, shown in Table 3.

Table 2. Frequency of Level 1 responses classified by linker type

Group	*de*	PP	RC	*con/sin*	Spatial
Children	499	23	613	18	5
Adult	55	8	63	5	3

Table 3. Frequency of recursive responses classified as linkers

Groups	*de*	PP	RC	*de*+RC	RC+*de*	RC+PP	*de*+PP	Other
Children	148	3	55	51	53	113	7	5
Adult	136	5	50	43	15	100	10	4

In the recursive responses, recursive *de* far outpaces recursive RCs at the ratio of 3:1. This is true of both children and adults. Children furthermore prefer RC+*de* in their mixed responses, which represented 12% of their target data, compared to 4% in adults. In a few cases, children used *de* to be able to link the prepositional phrase to the null nominal head.

(16) *el **de** encima del carro*   (01505NM, 5;02)
        "the one of on top of the car"

At Level 1 embedding, PPs are relatively rare, about 2% of the child data and 7% of the adult data. This is to be expected. However, among mixed target responses, the combination RC+PP is surprisingly high, making up 26% and 28% of the target data. A closer inspection suggests that these cases do not represent a true hypotactic configuration. Consider (17):

(17) a. *la que tiene el perro con sombrero*   (5AP 5;07)
            "The (one) that has the dog with hat"
      b. *el que tiene el vaso en la mesa*   (5CC 5;04)
            "the (one) that had the glass on the table"
      c. *de él que está debajo de la mesa junto a la ventana*
            "of he that is beneath the table next to the window"

Most of these examples involve *tener*, and are compatible with a small clause analysis without recursive embedding (*Tengo a Juan en la cocina* "I have John in the

kitchen"). Others contain a copula and are a possible case of ellipsis ("the one **that is** under the table ~~that is~~ next to the window"). We found eight PP+PP sequences, including three tokens produced by three different children and five produced by four adults. Three cases had recursive comitative *con* (one adult and one child (17)); the remainder contained mixed prepositions, including *con-sin, en/cerca*, and *de/en*.

(18) *la niña con el perro con sombrero* (6SM 6;03)
"The girl with the dog with hat"

Individually, the eight speakers producing iterated PPs were also the highest ranked producers of lexical prepositions at the single level. These children ranked 1st, 2nd and 4th out of 110 children in their frequency of use of lexical prepositions at Level 1; these adults were in the 1st, 3rd, 4th and 5th rank among 22 individuals in the sample. Despite how restricted PP modifiers are in the language, some speakers use them, and are able to produce recursive PP modifier sequences.

4.5 The bilingual study

One approach for testing the effect of experience is to consider alternative populations, such as bilinguals. The assumption is that bilingual children develop typically, while receiving quantitatively less exposure in one of their languages compared to monolingual peers. Input reduction has consequences: single-language vocabularies are on the average smaller in bilinguals compared to monolinguals (Pearson, Fernandez & Oller 1993). How bilingualism impacts the timing of grammatical development varies across domains. Many aspects of core sentence grammar develop with monolingual timing (Paradis & Genesee 1996), but other properties show delay (Pirvulescu et al.2014; Thordardottir 2014). Unsworth (2014) proposes that lexically-sensitive grammatical features are more likely to be delayed than formal properties.

This distinction relates to the second and third stages of DP acquisition. Level 1 modification (the second stage of development) involves the acquisition of the lexical linkers, and their lexical co-occurrence restrictions. When progressing to the recursive stage, the lexical ingredients have already been acquired. The child only needs to establish whether iteration is possible. Given our discussion above, such step might simply be an expression of productivity of the rule at level 1 and does not require additional lexical learning. If Unsworth (2014) is correct, and lexical sensitivity determines which structures are delayed in bilinguals, Level 1 embedding might be vulnerable in bilingual acquisition, while Level 2 embedding might be resilient. To investigate bilingual in the acquisition of recursive structures, Pérez-Leroux, Pettibone & Castilla-Earls (2017) compared production of recursive and non-recursive NPs with two modifiers (i.e., Level 2 vs. 2 Level 1 modifiers), as

in (19)–(20). Contrasting these structures serves to tease apart how bilingualism affects the (lexically-sensitive) rule acquisition stage vs. the rule-iteration stage.

(19)   Recursive modification (Level 2)
       [The dog [next to the tree [next to the house]]]

(20)   Sequential (Non-recursive) modification (2 Level 1)
       [The book [under the table] (and)] [in a box]]

Data from 35 simultaneous Spanish/English bilingual children from Toronto was compared to data taken from Pérez-Leroux et al. (2018) from 71 English monolingual children. Children were between the ages of 4 and 6, and the age difference between language groups was not significant. As in previous studies, both groups had fewer targets for the recursive condition than the non-recursive condition (Pérez-Leroux et al. 2018; Roberge et al. 2018). Figure 3 shows that the monolingual advantage was far greater for the non-recursive condition.
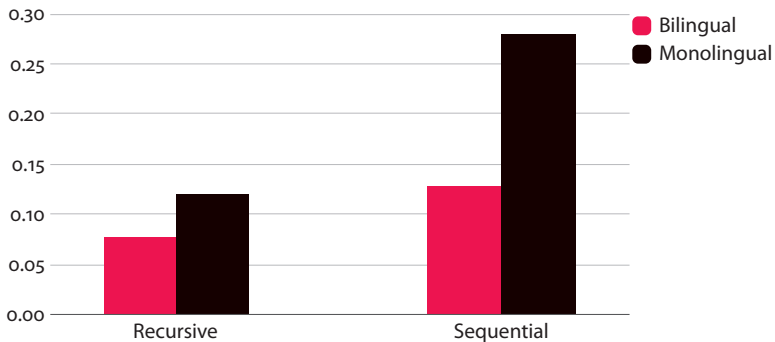
Figure 3. Average proportion of target responses to recursive and sequential double modification in English monolinguals and Spanish/English bilingual children (in English)

Individual analyses help explain the differences. When taken individually, more bilingual children were still at Stage 1 (single NP stage) compared to monolinguals (about 20% vs. 3%, respectively). Proportionally, more monolingual children were at Stage 2. Such children contribute target responses to the sequential condition while still failing in the recursive condition. The statistical analysis on the recursive condition showed that age, not bilingual status, was a significant determinant of whether a child was able to produce the target recursive responses. Since bilingualism impacts the number of bilingual children who have not yet progressed to the first level of embedding, we concluded that rule acquisition is vulnerable in bilingual acquisition. In contrast, the acquisition of rule iteration is a resilient property of grammar, less sensitive to fluctuations in the input.

5.   Discussion

What does the acquisition of recursion tell us about experience, and learning? In Spanish, as in other languages (Pérez-Leroux & Roberge 2018), recursive structures are present in some children at four, but become prevalent after age five. Recursion is not acquired simultaneously for all types of modifiers, even when the lexical elements are the same, as shown by the differences between possessives and relational nouns, both of which are expressed with *de* in Spanish. However, considering patterns of productivity across conditions, children follow the same path as adults.

De does not group with other Spanish prepositions. *De* phrases and relative clauses made up the overwhelming majority of modifiers. Speakers prefer *de* in contexts of recursive embedding; children do so even more than adults. Lexical PPs are rare generally, and even rarer is recursive embedding of lexical prepositions. Two speakers actually produced iterated embeddings of the same preposition, using the strict criteria, the same who produced most responses with the same prepositions at Level 1. In other words, speakers only use prepositions for recursive modification if they are frequent users of those prepositions for Level 1 embedding in the first place. This seemingly trivial point contains a non-trivial insight. Despite the low frequency of PP modifiers in Spanish and the range of usage restrictions, recursive embedding of PP modifiers can exist for speakers who use those PPs productively at the first level of embedding. This is not a formal demonstration of how productivity works in acquisition, but it suggests that the idea might be on the right track.

Assuming that productivity is a sufficient criterion for recursive embedding eliminates the poverty of stimulus problem, as the evidence to learn which phrases are recursive is contained in the distribution of Level 1 modifiers. The rare encounters with actual recursive structures in the input are not essential to progress to the third stage. The diachronic fluidity of Indo-European nominal recursion is clearly accounted for. Also explained is the observation that recursive structures are resilient in bilingual acquisition. Once the embedding rules are acquired, recursive modification catches up, and the monolingual's advantage is mitigated.

The debate between emergentists/functionalists and grammar-based/formalists has made little progress because the discussion is entangled in two non-overlapping senses of the word *learning*.[1] In developmental psychology, learning models the relationship between organism and the environment, with defined standards of evidence to prove that *x* causes *y* to learn *z* (training studies, cross-lagged correlation in cross-sectional studies, large scale longitudinal analyses, etc.). In the generative

---

1.   As pointed out by an anonymous reviewer, there is a third sense of the term, where *learning* refers to conscious mental activity and is contrasted to *acquisition*. Here we are only referencing implicit learning, i.e., acquisition.

literature, the term is used loosely to mean 'detection of properties', which are presumably matched against a set of pre-existing representations. It is time to abandon this perspective: "It doesn't have to be the way we thought it was" (Leguin 2017: 84).

Nothing is 'learned', except the linker that mediates between two nouns. This sits comfortably within the range of learning problems that the syntactic bootstrapping hypothesis was designed to solve. Learning psychology tells us that children are excellent distributional learners and generalizers, both domain-general capacities. UG enters the picture when the results of these learning generalizations are shown to be not random, nor dissociated from formal principles of structure, and in helping to understand the limits of possible combinations of properties and their relationships. Our results on the acquisition of recursive structure make most sense under assumptions that children have a rich implicit understanding of abstract structure, and are working with principles of computational efficiency. The input or evidence needed for learning recursive structures is distributional. This gives us evidence to support the position that learning (parametric properties) is based on "easily detectable properties" of language (Chomsky 2001: 2), and hence distributional, rather than identificational.

To the extent we see linguistics as part of the cognitive enterprise, our assumptions (in acquisition and theory) about the sphere of UG and the boundaries of domain specificity should be updated to accommodate a learning process that operates with domain-general tools but map instantaneously into a symbolic, highly constrained system. Only then can the gap between developmental and linguistically meaningful work be bridged. Important advances in the field are pushing parameters to more plausible levels of abstraction (see for instance, Longobardi 2018). Nonetheless, my critique of parameter setting as a 'learning' model still stands: there is no sense in which parameters contribute methods to update language specific grammatical representations. For recursive structures, learning the linker particle is a lexical problem, solved by lexical-distributional mechanisms (syntactic bootstrapping) and that is all there is. No parameter hides behind NN compounding in Spanish and English, or the Saxon genitive puzzle. We appear to have circled back to the traditional generative assumption of acquisition as rule-based. The difference is that we now know more about what rules can be, and how to deal with them. We know more about how the algorithmic, representational properties of language are implemented in computation and in biology. Children need learning mechanisms that are general enough to handle language variation and to explicitly address the challenges of integration across levels of implementation. These requirements were stated at the outset of the minimalist program in Chomsky (1995) and Hauser, Chomsky & Fitch (2002), and remain essential not only for explanatory adequacy, but for higher levels of adequacy, including typological, historical and evolutionary adequacy (Longobardi 2018).

Acknowledgements

Funding

References

Antonsen, Karoline & Romina Farhang. 2018. The effect of PP type on recursive modification. Poster presented at the meeting of the Canadian Linguistic Association, Regina, Saskatchewan, Canada, May 30-June 1.

Berko, Jean. 1958. The child's learning of English morphology. *WORD* 14(2–3). 150–177. https://doi.org/10.1080/00437956.1958.11659661

Brooks, Patricia J. & Michael Tomasello. 1999. How children constrain their argument structure constructions. *Language* 75(4). 720–38. https://doi.org/10.2307/417731

Brown, Roger & Jean Berko. 1960. Word association and the acquisition of grammar. *Child Development* 31(1). 1–14. https://doi.org/10.2307/1126377

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6. 339–405. https://doi.org/10.1023/A:1008324218506

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton. ISBN: 3-11-017279-8. https://doi.org/10.1515/9783112316009

Chomsky, Noam. 1959. A Review of B. F. Skinner's Verbal Behavior. *Language* 35(1). 26–58. https://doi.org/10.2307/411334

Chomsky, Noam. 1981. *Lectures on government and binding*. Holland: Foris Publications.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. Westport, CT: Praeger.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life inlLanguage*, 1–52. Cambridge, MA: MIT Press.

Clark, Eve. 1995. *The lexicon in acquisition*. Cambridge: Cambridge University Press.

Eisenberg, Sarita, Teresa Ukrainetz, Jennifer Hsu, Joan Kaderavek, Laura Justice, & Ronald Gillam. 2008. Noun phrase elaboration in children's spoken stories. *Language, Speech, and Hearing Services in Schools* 39. 145–157. https://doi.org/10.1044/0161-1461(2008/014)

Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals. *Behavioral and Brain Sciences* 32. 429–492. https://doi.org/10.1017/S0140525X0999094X

Everett, Daniel L. 2005. Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology* 46. 621–46. https://doi.org/10.1086/431525

Fitch, W. Tecumseh. 2010. *The evolution of language*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511817779

Fodor, Janet. 1998. Unambiguous triggers. *Linguistic Inquiry* 29. 1–36. https://doi.org/10.1162/002438998553644

Futrell, Richard, Laura Stearns, Daniel L. Everett, Steven T. Piantadosi & Edward Gibson. 2016. A corpus investigation of syntactic embedding in Pirahã. *PLOS ONE* 11. e0145289. https://doi.org/10.1371/journal.pone.0145289

Gleitman, Lila R. 1990. The structural sources of verb meaning. *Language Acquisition* 1 (1). 3–55. https://doi.org/10.1207/s15327817la0101_2

Gleitman, Lila R., Kimberly Cassidy, Rebbeca Nappa, Anna Papafragou & John C. Trueswell. 2005. Hard words. *Language learning and development* 1(1). 23–64. https://doi.org/10.1207/s15473341lld0101_4

Goldin-Meadow, Susan. 1982. The resilience of recursion. In Eric Wanner & Lila R. Gleitman (eds.), *Language acquisition: The state of the art*, 51–77. New York: Cambridge University Press.

Guasti, Maria Teresa, Anna Gavarró, Joke de Lange & Claudia Caprin. 2008. Article omission across child languages. *Language Acquisition* 15(2). 89–119. https://doi.org/10.1080/10489220801937346

Han, Chung-hye, Jeffrey Lidz & Julien Musolino. 2016. Endogenous sources of variation in language acquisition. *Proceedings of the National Academy of Sciences* 113(4). 942–947. https://doi.org/10.1073/pnas.1517094113

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298. 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Heim, Irene. 1991. Artikel und Definitheit [Articles and definiteness]. In Armin von Stechow & Dieter Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, 487–535. Berlin: de Gruyter. https://doi.org/10.1515/9783110126969.7.487

Hunsicker, Dea & Susan Goldin-Meadow. 2012. Hierarchical structure in a self-created communication system. *Language* 88. 732–63. https://doi.org/10.1353/lan.2012.0092

Leguin, Ursula. 2017. *No time to spare*. Boston: Houghton Mifflin Harcourt.

Lidz, Jeff, Susan Waxman & Jennifer Freedman. 2003. What infants know about syntax but couldn't have learned. *Cognition* 89(3). B65–B73. https://doi.org/10.1016/S0010-0277(03)00116-1

Lightfoot, David. 1991. *How to set parameters: Arguments from language change*. Cambridge, MA: MIT Press.

Limbach, Maxi & Dani Adone. 2010. Language acquisition of recursive possessives in English. *Proceedings of the Boston University Conference on Language Development* 34. 281–90.

Longobardi, Giuseppe. 2018. Principles, parameters, and schemata: A radically underspecified UG. *Linguistic Analysis* 41 (3). 517–557.

Marcus, Gary F., Sugumaran Vijayan, S. Bandi Rao & Peter M. Vishton. 1999. Rule learning by seven-month-old infants. *Science* 283(5398). 77–80. https://doi.org/10.1126/science.283.5398.77

Markman, Ellen M. & Gwin F. Wachtel. 1988. Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology* 20. 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Moreira-Rodríguez, Antonia. 2006. 'The book *on the table*', 'The nan *in the moon*': Post-modification of nouns by preposition + noun in English and Castilian. *Bulletin of Spanish Studies* 83(1). 53–72. https://doi.org/10.1080/1475382062000346045

Nevins, Andrew, David Pesetsky & Cilene Rodrigues. 2009. Pirahã exceptionality: A reassessment. *Language* 85. 355–404. https://doi.org/10.1353/lan.0.0107

Ninio, Anat. 2011. *Syntactic development, its input and output*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199565962.001.0001

Pannemann, Maren. 2007. DP acquisition as structure unravelling. Amsterdam: Universiteit van Amsterdam dissertation.

Paradis, Johanne & Fred Genesee. 1996. Syntactic acquisition in bilingual children. *Studies in Second Language Acquisition* 18 (1). 1–25. https://doi.org/10.1017/S0272263100014662

Pearl, Lisa & Jeff Lidz. 2013. Parameters in language acquisition. In Cedruc Boeckx & Kleanthes K. Grohmann (eds.), *Cambridge handbook of biolinguistics*, 129–159. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511980435.010

Pearson, Barbara Zurer, Sylvia C. Fernandez, & D. Kimbrough Oller. 1993. Lexical development in bilingual infants and toddlers. *Language Learning* 43 (1). 93–120. https://doi.org/10.1111/j.1467-1770.1993.tb00174.x

Pérez-Leroux, Ana T., & Anny P. Castilla-Earls. 2016. Adquisición del español como lengua materna. In Javier Gutierrez-Rexach (ed.), *Enciclopedia de lingüística hispánica*, 213–230. New York: Routledge. https://doi.org/10.4324/9781315713441-93

Pérez-Leroux, Ana T., Anny Castilla-Earls, Susana Béjar & Diane Massam. 2012. Elmo's sister's ball: The problem of acquiring nominal recursion. *Language Acquisition* 19. 301–11. https://doi.org/10.1080/10489223.2012.685019

Pérez-Leroux, Ana T., Tyler Peterson, Anny Castilla-Earls, Susana Béjar, Diane Massam & Yves Roberge. 2018. The acquisition of recursive modification in NPs. *Language* 94. 332–359. https://doi.org/10.1353/lan.2018.0019

Pérez-Leroux, Ana T., Erin Pettibone & Anny Castilla-Earls. 2017. Down two steps: Are bilinguals delayed in the acquisition of recursively embedded PPs? *Matraga (Aquisição da Linguagem)* 24(41). 393–416. https://doi.org/10.12957/matraga.2017.28781

Pérez-Leroux, Ana T. & Yves Roberge. 2014. *Development of noun phrase complexity in children*. Social Sciences and Humanities Research Council of Canada. Insight Grant Program. SSHRC: 435-2014-2000.

Pérez-Leroux, Ana T., Yves Roberge, Alex Lowles, & Petra Schulz. 2021. Structural diversity does not affect the acquisition of recursive embedding: The case of possession in German. *Language Acquisition*. doi.org/10.1080/10489223.2021.1965606

Pérez-Leroux, Ana T., Castilla-Earls, María Fernanda Lara Díaz, & Erin Pettibone. In preparation. *La adquisición de estructuras nominales complejas*. University of Toronto.

Pesetsky, David. 2009. Against taking linguistic diversity at face value. *Behavioral and Brain Sciences* 32(5). 462–465. https://doi.org/10.1017/S0140525X09990562

Pettibone, Erin. In preparation. The acquisition of (non)restrictive adjectives. Toronto, Canada: University of Toronto dissertation.

Picallo, Carme. 2012. Structure of the noun phrase. In Jose Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds), *The handbook of Hispanic linguistics*, 263–284. Chichester, UK: John Wiley & Sons. https://doi.org/10.1002/9781118228098.ch14

Pine, Julien M. & Elena V. M. Lieven. 1997. Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics* 18. 123–138. https://doi.org/10.1017/S0142716400009930

Pirvulescu, Mihaela, Yves Roberge, Danielle Thomas, Ana T. Pérez- Leroux, & Nelleke Strik. 2014. Bilingual effects: Exploring object omission in pronominal languages. *Bilingualism: Language and Cognition* 17(3). 495–510. https://doi.org/10.1017/S1366728913000631

Pinker, Stephen. 1984. *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Pérez-Leroux, Ana T. & Yves Roberge. 2018. A way into recursion. In Hollebrandse, Bart, Jaieun Kim, Ana T. Pérez-Leroux, & Petra Schulz (Eds.), *UMOP 41: Thoughts on Mind and Grammar (T.O.M. and grammar): A Festschrift in Honor of Tom Roeper*, 109–122. Amherst, MA: GLSA.

Roberge, Yves, Pérez-Leroux, Ana T. & Anna Frolova. 2018. On recursive modification in L1 French. *Languages* 3 (6). https://doi.org/10.3390/languages3010006

Roeper, Tom. 2011. The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics* 5. 57–86

Roeper, Tom & William Snyder. 2005. Language learnability and the forms of recursion. In Anna Maria Di Sciullo (ed.), *UG and external systems: Language, brain and computation*, 155–169. Amsterdam: John Benjamins. https://doi.org/10.1075/la.75.10roe

Roeper, Tom & Jill G. de Villiers. 2011. The acquisition path for *wh*-questions. In Jill G. de Villiers & Tom Roeper (eds.), *Handbook of generative approaches to language acquisition*, 189–246. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-1688-9_6

Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month old infants. *Science* 274. 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Sauerland, Uli. 2006. On the semantic markedness of phi features. In Daniel Habour, David Adger & Susana Béjar (eds.), *Phi-theory: Phi features across modules and interfaces*, 57–82. Oxford University Press.

Sauerland, Uli & Hans-Martin Gärtner. 2007. *Interfaces + recursion = language?* Berlin: Mouton de Gruyter.

Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. Paris: Ed. Payot.

Sneed German, Elisa. 2008. Input in the acquisition of genericity. Evanston, IL: Northwestern University dissertation.

Snyder, William. 2007. *Child language: The parametric approach*. Oxford, UK: Oxford University Press.

Snyder, William, Ann Senghas & Kelly Ihnman. 2001. Agreement morphology and the acquisition of noun-drop in Spanish. *Language Acquisition* 9(2).157–173. https://doi.org/10.1207/S15327817LA0902_02

Syrett, Kristen & Jeff Lidz. 2009. QR in child grammar: Evidence from antecedent-contained deletion. *Language Acquisition* 16. 67–81. https://doi.org/10.1080/10489220902769226

Thordardottir, Elin. 2014. The typical development of simultaneous bilinguals. In Johanne Paradis and Therese Grüter (eds.), *Input and experience in bilingual development*, 141–160. Amsterdam: John Benjamins. https://doi.org/10.1075/tilar.13.08tho

Tomasello, Michael. 2000. Do young children have adult syntactic competence? *Cognition* 74. 209–253. https://doi.org/10.1016/S0010-0277(99)00069-4

Unsworth, Sharon. 2014. Comparing the role of input in bilingual acquisition across domains. In Johanne Paradis & Therese Grüter (eds.), *Input and experience in bilingual development*, 181–201. Amsterdam: John Benjamins. https://doi.org/10.1075/tilar.13.10uns

Valian, Virginia, Stephanie Solt & John Stewart. 2009. Abstract categories or limited-scope for-
mulae: The case of children's determiners. *Journal of Child Language* 36(4). 743–778.
https://doi.org/10.1017/S0305000908009082

Widmer, Manuel, Sandra Auderset, Johanna Nichols, Paul Widmer & Balthasar Bickel. 2017. NP
recursion over time: Evidence from Indo-European. *Language* 93. 799–826.
https://doi.org/10.1353/lan.2017.0058

Yang, Charles. 2002. *Knowledge and learning in natural language*. New York: Oxford University
Press.

Yang, Charles. 2004. Universal Grammar, statistics, or both. *Trends in Cognitive Science* 8(10).
449–454. https://doi.org/10.1016/j.tics.2004.08.006

Yang, Charles. 2015. For and against frequencies. *Journal of Child Language* 42(2). 287–293.
https://doi.org/10.1017/S0305000914000683

Yang, Charles. 2016. *The price of linguistic productivity*. Cambridge, MA: MIT Press.
https://doi.org/10.7551/mitpress/9780262035323.001.0001

Definite determiners in Romance

The role of modification

Daniela Isac

Concordia University

The definite article in Eastern Romance (ER) is overtly expressed as a suffix on the noun. However, with objects of prepositions (Ps) the definite article can be null, unless the P object is modified, in which case the definite article must be overt. In order to account for this variation in the overtness of the definite article, I propose that definite Ds must undergo M-merge in ER and that the definite D is spelled out overtly only if D M-merges with a head that bears a definiteness ([def]) feature. ER languages display micro-variation with respect to which particular item in the DP can host the definite article, and with respect to whether the definite article in unmodified objects of Ps must, or simply can, be null. Micro-variation is accounted for by (i) the feature content of the Number head, and (ii) whether the M-merge rule applies strictly locally or not.

1.   Introduction

This chapter focuses on one property of definite DPs that can be observed in Eastern Romance (ER) languages (Romanian – R, Northern Aromanian – NAr, Southern Aromanian – SAr, Megleno-Romanian – MR, Istro Romanian – IR), but not in other Romance languages, like Spanish, Italian, or French. While the definite D in ER is normally overt, with objects of Ps in the Accusative case, the definite D can be null. However, if the object of P is modified, the definite D must be overt. All the examples in this article include adjectival (A) modifiers, but modifiers that are PPs or relative clauses behave similarly.

(1) a. *în parc(-*ul)*   (R)
         in park-DEF
         "in the park"

b.   *în parc-ul   cu   tei.*   (R)
      in park-DEF with linden
      'in the park with linden trees'

Although this property is shared by all ER languages, micro-variation can be observed among these languages. For example, in R the definite D of an unmodified object of P must be null, but in the other ER languages it can, but does not have to, be null. Moreover, R does not allow the double expression of the definite article on both N and A, but all the other ER languages do allow it. The tables below sum up the data that is the focus of this chapter. Concrete examples from each language will be included in the relevant sections in the chapter. Table 1 illustrates the overtness of the definite article on the N with simple objects of Ps, Table 2 includes DPs with postnominal As, and Table 3 shows the overt expression of the definite article in DPs with prenominal As.

Table 1. P-N

	P-N$_{def}$	P-N
R	*	✓
NAr	✓	✓
SAr	✓	✓
IR	✓	✓
MR	✓	✓

Table 2. P-N-A

	P-N$_{def}$-A	P-N-A$_{def}$	P-N$_{def}$-A$_{def}$
R	✓	*	*
NAr	✓	*	✓
SAr	*	*	✓
IR	✓	*	✓
MR	✓	*	✓

Table 3. P-A-N

	P-A-N$_{def}$	P-A$_{def}$-N	P-A$_{def}$-N$_{def}$
R	*	✓	*
NAr	*	✓	*
SAr	*	*	*
IR	*	✓	✓
MR	*	✓	✓

The main questions addressed in this chapter are the following: (i) how can we account for the two properties that all ER languages share, i.e. (a) the fact that the definite article can be null with unmodified objects of Ps; and (b) the fact that the definite article must be expressed overtly with modified objects of Ps (either on the N, or on the A modifier, or on both); (ii) how can we account for the variation captured in the tables above with respect to which item in the DP can host the definite article across ER?

2.   Existing literature

The existing generative literature on definiteness in ER languages is focused mainly on DPs in non-prepositional environments, with the exception of Mardale (2006). In Mardale's view, the definite D of objects of Ps incorporates into P, a process which is made possible by the reduced syntactic structure of these DPs. In particular, objects of Ps are assumed to lack a K(ase)P and a NumberP.

(2) $[_{PP} P[_{(KP)} (K) [_{DP} D [_{(NumP)} (Num) [_{NP} N]]]]]$

Since Mardale (2006) assumes Dobrovie Sorin & Giurgea's (2006) analysis of definiteness in Romanian, according to which D lowers to Num and gets spelled out on the N which has raised to Num, the absence of Num guarantees that D will not get spelled out on N, but will incorporate into P.

   Apart from the fact that the posited syntactic defectiveness of objects of Ps is somehow stipulative, Mardale's (2006) analysis cannot really account for the role of modification in licensing the overt definite D. Mardale proposes that incorporation of D into P is blocked by the modifier because the latter forces the projection of a DP, which in turn blocks incorporation. However, the DP must be projected even in the absence of a modifier, since it is D that incorporates into P, and so it is not clear why there should be a difference between the DP layer for a modified object and the DP layer of a non-modified one.

3.   Theoretical background

The present analysis relies on Pesetsky & Torrego's (2007) view on feature valuation, as well as on a number of assumptions about the syntax of DPs that are detailed below.

3.1   Features

Pesetsky & Torrego (2007) distinguish between interpretable/ uninterpretable features on the one hand, and valued/unvalued features on the other, and propose a fourfold feature typology: (a) uninterpretable, valued features [uF:val]; (b) interpretable, valued features [iF:val]; (c) uninterpretable, unvalued features [uF:]; (d) interpretable, unvalued features [iF:]. There are two important differences between this approach and Chomsky's (2000, 2001) theory of features. First, in Pesetsky & Torrego's (2007) approach, Agree is always initiated by a head with an unvalued feature (uninterpretable or interpretable), as opposed to being driven by the need to delete uninterpretable features. Second, Agree is conceived of as feature sharing in Pesetsky & Torrego's approach. What this means is that once a probe finds a goal with a matching feature, a link/chain is created between positions which have agreed, which is accessible throughout the derivation. So, if a feature F on a head X is still unvalued after a first search, X will reinitiate the search and another operation of Agree will apply if X finds a goal. Crucially, if the F that is found is valued, then feature sharing will result in a valued feature F present at three locations:

(3) $(X_{[F:]}, Y_{[F:]}) \ldots Z_{[F:val]}$--Agree$\rightarrow$($X_{[F:val]}$, $Y_{[F:val]}$, $Z_{[F:val]}$)

In applying this system to the features of the DP, I follow Cornilescu & Nicolae (2011) and assume that Romanian Ns bear a valued [def] feature and that other items within the DP that have a [def] feature (including Ns and As) have an unvalued instance of this feature. I extend this view to the other ER languages given that all these languages have a suffixal definite article. Given its unvalued [def] feature, D searches for a matching feature in its c-command domain and establishes a link with its goals, potentially with both A and N. In the rest of the chapter I will refer to the set of heads that are linked by sharing a [def] feature a definiteness valuation chain.

3.2   Prepositions

I follow Pesetsky & Torrego (2004) in assuming that Ps are merged DP internally in a position that is analogous to the position occupied by T within the CP. The reasoning behind this view on Ps has to do with Case. In Pesetsky & Torrego's view all instances of structural case are instances of [uT] on D and with objects of Ps it is the P that values the [uT], i.e. Case, feature on D. Moreover, P also gets attracted to D by the latter's EPP feature. Given that in Pesetsky & Torrego's view head movement and phrasal movement both target the specifier position of the attracting head, when the P head is attracted by D, it moves to Spec,D, as shown in (4).

(4) $[_{DP}$ P $[_{D'}$ D $[_{PP}$ ~~P~~ NumP]]]

On the other hand, if the P is too distant from D, what gets attracted to Spec,D is not the P head, but the whole PP. Thus, the choice between P and PP movement is a function of the syntactic distance between P and D.

(5)   Pesetsky & Torrego (2001, 2007)
       if a head H attracts a feature of XP as part of a movement operation, then
       a.   if XP is the complement of H, copy the head of XP into the local domain of H
       b.   otherwise, copy XP into the local domain of H

3.3   Nominal phases and their peripheries

I follow Cornilescu & Nicolae (2011); Tănase Dogaru (2012), among others, in assuming that DPs consist of two phases –NumP and DP, roughly equivalent to the vP and CP in the clausal domain. Each of these two phases have a left periphery, including projections where topic and focus material might get dislocated in the course of the derivation. I will use the periphery features proposed by Lopez (2009): [+a(naphoric)], which expresses an obligatory link to an antecedent, and [+c(ontrast)], which identifies a referent in relation to a set of alternatives, by excluding the other alternatives.

3.4   The number phrase

Given that NumP bears an [N] feature, as well as periphery features like [c] or [a] in all ER languages, the following feature combinations are possible: [N] and [a] (which yields anaphoric interpretations), [N] and [c] (which results in contrastive interpretations), [N] and both [a] and [c] (for contrastive topic interpretations). Moreover, the EPP could be in principle associated with any of these features. The total number of possible feature specifications given these assumptions is seven but I propose that some feature sets are ruled out by the following restriction: [a] is always associated with an EPP in ER. The remaining feature matrices are given below:

(6) Num_1: [uN], [ua:]/EPP
       Num_2: [uN], [ua:]/EPP, [uc:+]
       Num_3: [uN], [uc:+]/EPP
       Num_4: [uN]/EPP, [uc:+]

There are two features of the Num head that have a syntactic impact. First, the feature associated with an EPP will trigger the dislocation of an XP with a matching

feature to Spec,Num. Depending on which feature bears the EPP, various types of XPs will raise to Spec,Num across ER. Second, the presence of an [uc:+] feature on Num has consequences for the syntax of the higher projections in the DP. A NumP headed by a Num head marked as [uc:+] will raise to the left periphery of the DP phase, in order to check the unvalued [c] feature on a Contrastive (Contr) head.

3.5   Adjectives

The literature describes two modes in which As can combine with nominals: functional application and predicate modification. Functional application combines constituents that have denotations of different types (Heim & Kratzer 1998): one constituent represents the function that applies to the other constituent, i.e. the argument. Predicate modification, on the other hand, combines constituents of the same denotational type (Higginbotham 1985); both constituents in this case are predicates. The syntactic correlate of a predicate modification relation is an adjunction configuration, while function application is syntactically mediated a functional head. Given that definite nominals in ER bear a [def] feature valued as '+', the respective NP is semantically an individual. Hence, the mode of combination with an AP in ER must always be function application, since As are predicates and the nominals are individuals (i.e. the two are of different types). For postnominal As I will therefore follow Kayne (1994); Alexiadou (2001); Cinque (2010), among others and propose that they have a predicative source in a reduced relative clause.

(7)   [$_{DP}$ D [$_{NumP}$ Num [$_{CP}$ NP [$_{C'}$ C [$_{IP}$ ~~NP~~ AP]]]]]

Given that postnominal APs are restrictive (Teodorescu 2006; Cornilescu 2004; Cinque 2010; Cornilescu & Dinu 2013), they introduce a contrast between the modified object and other NPs of the same type, that do not have the property denoted by the adjective. I will thus assume that the C head of the reduced relative clause bears a contrastive [c] feature. This feature triggers movement of the CP to a left periphery position–ContrP.

(8)   [$_{DP}$ D [$_{ContrP}$ CP [$_{Contr'}$ Contr [$_{NumP}$ Num [$_{\sim CP}$ NP [$_{C'}$ C [$_{IP}$ ~~NP~~ AP]]]]]]]

Prenominal As are similar to postnominal ones in that both types of As denote predicates which apply to the individual level denotation of the definite N(P). However, in order to capture the syntactic differences between the two, I follow Cornilescu (2004); Cinque (2010), and others in proposing that prenominal adjectives are merged as specifiers of functional projections rather than as complements.

(9)   [$_{NumP}$ Num [$_{FP}$ AP [$_{F'}$ F NP]]]

For the purposes of this chapter I disregard the semantic differences between various subtypes of prenominal As, and implicitly questions regarding the relative ordering of these As. What is important for the analysis proposed here is that all of these subtypes of prenominal As share one semantic property, they are all modal and quantificational (Cornilescu 2004; Cornilescu & Dinu 2013; Cinque 2010; Teodorescu 2006, among others). I similarly propose that F, the functional head that licenses these As, bears a quantificational feature which I take to be contrastive, given that prenominal As express context bound properties of definite NPs with identified referents. Notice that in the analysis proposed here, both heads that license As (C and F) have a [c] feature, but the A occupies different positions with respect to the licensing head (the A is in the Spec of the respective head if prenominal, and it is in a complement position if postnominal). The syntactic consequence of both C and F bearing a [c] feature is that both the CP and the FP must raise to the Spec,Contr to check their [c] feature.

4. Proposal

4.1 M-merge

In order to account for the properties of definite Ds in ER, I build on Matushansky's (2006) proposal that heads that move to the Spec of a head H can further undergo M-merge with H itself. According to Matushansky (2006) and Pesetsky & Torrego (2004), M-merge is triggered by some attracting heads but not all. For Pesetsky & Torrego (2004), who discuss English Ps, P movement to Spec,D is not followed by M-merge. I propose instead that in ER the P can M-merge with definite D.

(10)
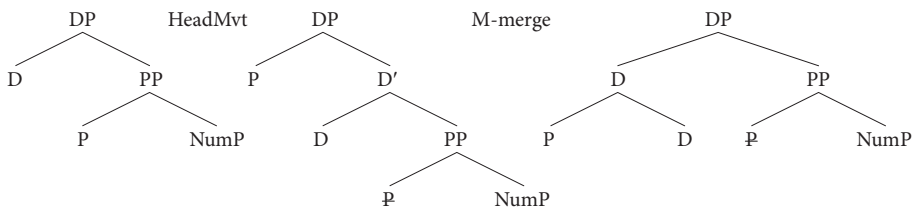
Moreover, I propose that M-merge is a consistent property of definite Ds in ER, which applies even in the absence of P heads moving to Spec,D.

(11) Definite Ds must undergo M-merge in ER.
    i.   If a head moves to its Spec, definite D M-merges with the respective head.
    ii. Otherwise, M-merge will target the heads that are part of the definiteness valuation chain.

This rule applies to definite Ds in all ER languages, and is related to the affixal status of definite Ds in these languages. While the rule in (11) applies to all ER languages, the exact head(s) that the definite D will M-merge with in a particular language depends on whether the respective language has a strictly local version of the rule or allows the 'spread' of the definite article to all the heads in the valuation chain. Thus, (11) can be rephrased as follows:

(12) Definite Ds must undergo M-merge in Eastern Romance.
     i.   If a head moves to its Spec, definite D M-merges with the respective head.
     ii. Otherwise, M-merge will target the heads that are part of the definiteness valuation chain.
          a.   in languages where M-merge is strictly local, definite D M-merges only with the closed head in the definiteness valuation chain
          b.   in languages that allow the spread of M-merge, definite D M-merges with all the heads in its definiteness valuation chain

4.2   Spell-out

I assume, together with Cornilescu & Nicolae (2009, 2011) and Dobrovie-Sorin & Giurgea (2006), among others, that the overtness of the definite D is not a consequence of N-to-D movement but a PF phenomenon. More specifically, the definite article is not necessarily spelled out in the D head per se. In fact, the definite D head is always covert in ER and the definite article is always spelled out elsewhere (on the N, or on a prenominal A or cardinal). For intuitive reasons, I will assume that only heads that bear a [def:+] feature are able to host the overt definite article, and I propose that the following Spell-out rule for ER:

(13) **Spell-out:** the definite D gets spelled out on the head(s) that it M-merges with, iff that head/those heads bears a [def] feature. Otherwise, the definite D will remain covert.[1]

---

1.   The above proposal on the Spell-out of the definite article is similar in spirit to the rule proposed by Dobrovie Sorin & Giurgea (2006) for Romanian. However, unlike Dobrovie Sorin & Giurgea's D lowering rule, which is sensitive to syntactic structure, the rule in (13) is based on the linear order of heads at PF, rather than on syntactic structure. This allows Ds to M-merge with the head of an XP in the Spec of its complement, for example, rather than with the head of its complement.

    The rule in (13) is also similar to Cornilescu & Nicolae's (2011) proposal that the definite article is pronounced at PF on the highest item below D bearing a [def] feature, on condition this item has nominal features (i.e. N or A). The differences between the rule in (13) and this proposal have to do with M-merger, an operation that is part of the present proposal but not Cornilescu & Nicolae's. What M-merger achieves for the analysis in this paper is the possibility of accounting for the (lack of) Spell-out of definite articles with objects of Ps.

Notice that there are three potential outputs of the Spell-out rule: (i) the definite D can be covert (if a head moves to Spec,D); (ii) for languages in which M-Merge is strictly local, the definite D can be overt only on one item (i.e. the head that is part of the definiteness valuation chain that is closest to D); and (iii) in languages in which M-Merge can spread, the definite D can be overt on multiple items, i.e. on all the heads in the definiteness valuation chain.
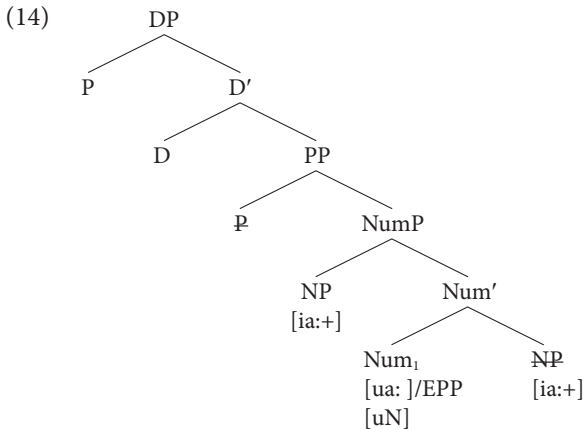
5.   Implementation of the analysis

5.1   Num heads and the expression of definiteness

Depending on the exact features that a Num head bears, i.e. which periphery features it bears and which of its features is marked as EPP, the ordering of sub-constituents within the DP will be different. These different orderings will in turn interact with the M-merge and Spell-out rules proposed above and will result in the definiteness feature overtly expressed on various items. Given the possible features on Num discussed in § 3.4 and shown in (6), the following derivations could take place:

5.1.1   *Num$_1$: [uN],[ua:]/EPP*
If Num has the features of Num$_1$, the structure is as in (14).

(14)

```
                      DP
                    /    \
                   P      D'
                        /    \
                       D      PP
                            /    \
                           P̶    NumP
                                /    \
                             NP      Num'
                           [ia:+]   /    \
                                 Num₁     N̶P̶
                              [ua: ]/EPP  [ia:+]
                                [uN]
```

A Num head marked as [a] can result in a grammatical DP string only if the structure contains no A. This is because As are licensed by a functional head that bears an uninterpretable [c] feature (either the C head, for postnominal As, or F, for prenominal As). The only way in which the uninterpretable [c] feature on these heads can be checked is if the CP/FP is attracted to Spec,Num, i.e. to the edge of the NumP phase. Otherwise, the [c] feature on these heads cannot be 'seen' by a Contr

head merged above NumP and searching for a matching feature. In a structure like (14), on the other hand, the NP which bears an [a] feature will raise to Spec,Num to check the EPP on the [a] feature of Num. The P head raises to Spec,D and subsequently M-merges with the latter. This prevents the definite article from getting spelled out, according to the rules in (12) and (13), and the only string that can be generated from (14) is P-N.

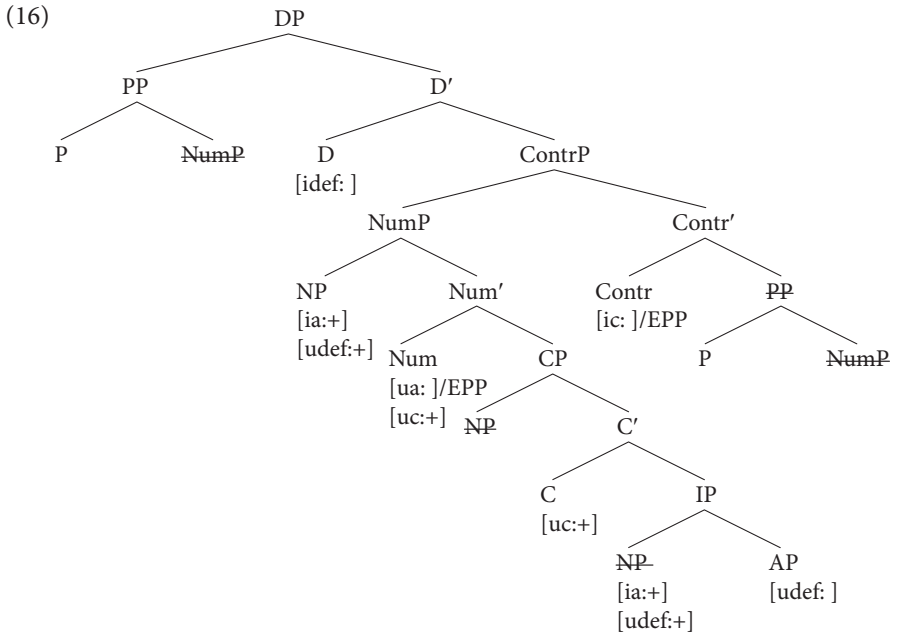### 5.1.2   *Num$_2$: [uN],[ua:]/EPP,[uc:+]*

If the Num head has both an [a] feature (marked as EPP) and a [c] feature, Num will attract an item with a matching [a] feature, just as in the case of Num$_1$. However, NumP will further be attracted to ContrP due to its [c] feature. Depending on whether or not the string contains an A, and on whether A is pre- or postnominal, three structures can obtain. If there is no A, the structure is as in (15). In (15) the NP raises to Spec,Num to check the [a]/EPP feature on this head and the NumP further raises to ContrP to check the [c] feature. Given that NumP raises to Spec,Contr, the P cannot move as a head to SpecD, since P is no longer the complement of D. Instead, the whole (remnant) PP raises to Spec,D, the definite D M-merges with the only head in its definiteness valuation chain, i.e. N, and the [def] feature is spelled out on N. The string that results is P-Ndef.

(15)
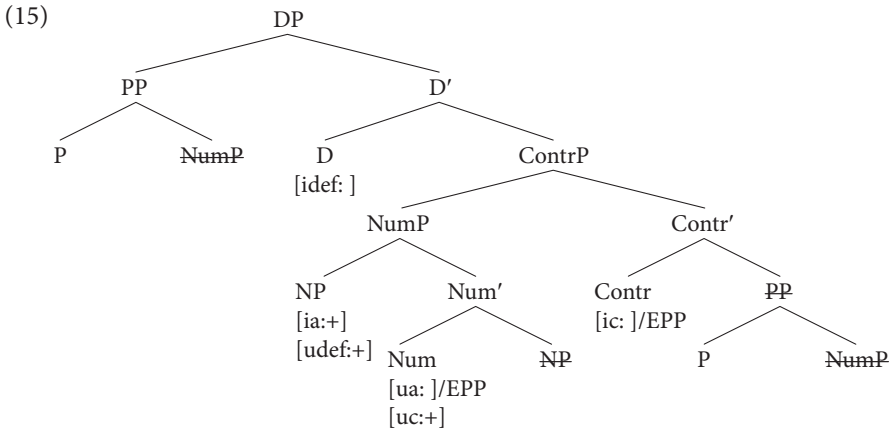
If a postnominal A is present, the structure is as in (16). Num again attracts NP, by virtue of its [a]/EPP feature, and the whole NumP raises to Spec,Contr to check the [c] feature. As a result, P is prevented from raising to Spec,D as a head, and the whole (remnant) PP moves to Spec,D. The definite D will M-merge with one or all the heads in its definiteness valuation chain, depending on whether M-merge is strictly local or it allows spread in the respective language.

(16)

```
                            DP
              ┌──────────────┴──────────────┐
             PP                            D′
          ┌───┴───┐              ┌──────────┴──────────┐
          P     NumP             D                   ContrP
                             [idef: ]      ┌───────────┴───────────┐
                                         NumP                    Contr′
                                  ┌────────┴────────┐      ┌────────┴────────┐
                                 NP              Num′    Contr              PP
                               [ia:+]         ┌───┴───┐  [ic: ]/EPP    ┌─────┴─────┐
                               [udef:+]      Num     CP                 P        NumP
                                          [ua: ]/EPP ┌──┴──┐
                                          [uc:+]    NP    C′
                                                       ┌───┴───┐
                                                       C      IP
                                                     [uc:+]  ┌──┴──┐
                                                            NP    AP
                                                          [ia:+] [udef: ]
                                                          [udef:+]
```

Notice that in this structure the definite D can only be spelled out on N, given that the NumP is a phase and the NP is the only material that is at the edge of NumP and accessible for Agree with D. The A on the other hand is too deeply embedded inside the NumP and inaccessible for Agree with D. The only output of a structure like (16) is thus a P-N$_{def}$-A string, and this is regardless of whether M-merge is strictly local or it allows spread.

Finally, if a prenominal A is present, the structure is as in (17).

(17)

```
                            DP
              ┌──────────────┴──────────────┐
             PP                            D′
          ┌───┴───┐              ┌──────────┴──────────┐
          P     NumP             D                   ContrP
                             [idef: ]      ┌───────────┴───────────┐
                                         NumP                    Contr′
                                  ┌────────┴────────┐      ┌────────┴────────┐
                                 AP              Num′    Contr              PP
                               [ua: ]         ┌───┴───┐  [ic: ]/EPP    ┌─────┴─────┐
                               [udef:+]      Num     FP                 P        NumP
                                          [ua: ]/EPP ┌──┴──┐
                                          [uc:+]    AP    F′
                                                  [ua: ] ┌───┴───┐
                                                  [udef: ] F     NP
                                                         [uc:+] [udef:+]
```
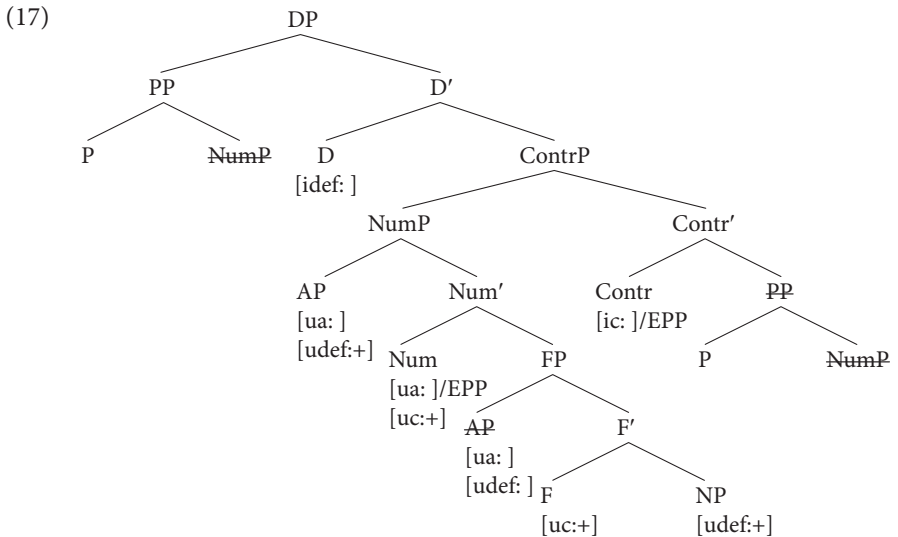
In (17) Num attracts AP, rather than NP, to its Spec, given that AP is the closest item with an [a] feature. NumP will raise to Spec,Contr to check its [c] feature and will consequently land between D and P. Hence P will be unable to move as a head, and the whole (remnant) PP will raise to Spec,D. According to the M-merge and Spell-out rules above, the definite D will M-merge with one or all the heads in its definiteness valuation chain. The only head that is accessible for Agree when D searches is A, as A is at the edge of the NumP phase. Since A is the only head in D's definiteness valuation chain, the definite D is spelled out on A only, regardless of whether M-merge is local or it allows spread. The only possible output of (17) is P-A_{def}-N.

5.1.3   *Num₃: [uN],[uc:+]/EPP*

A Num head with a [c]/ EPP feature can be merged only if a constituent with a matching [c] feature is present; otherwise the derivation will crash. A Num₃ head is thus grammatical only in a sequence that contains an A (either prenominal or postnominal), as As are licensed by heads that bear a [c] feature (the C head, for postnominal As, or the F head, for prenominal As). If a postnominal A is merged, the structure is as in (18).
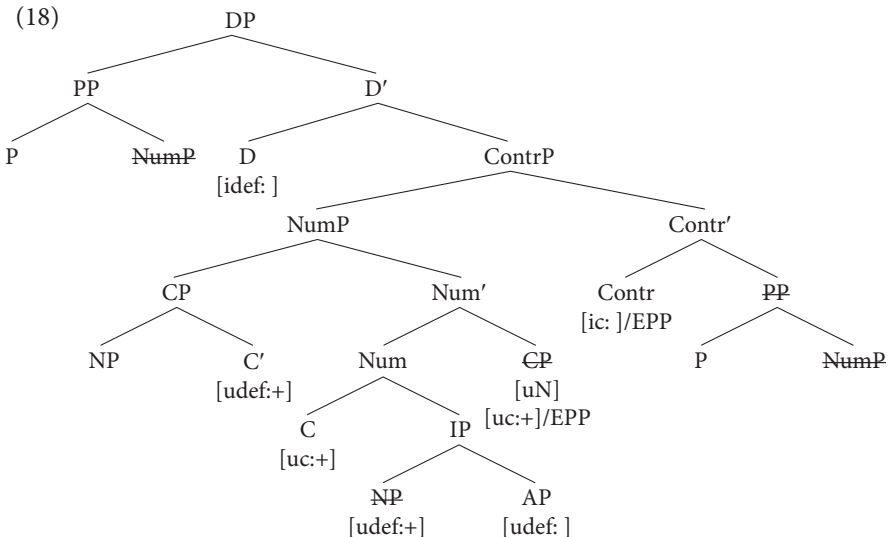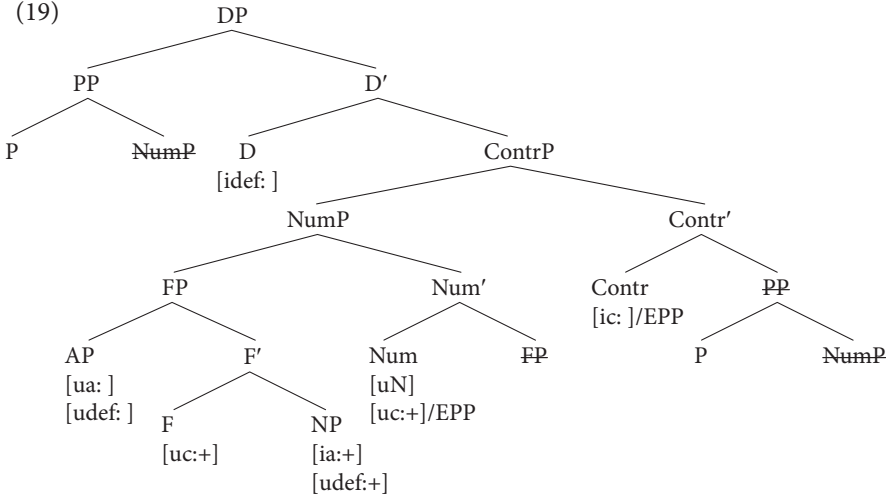

In (18) Num attracts the CP to its Spec and the whole NumP raises to Spec,Contr to check the [c]/EPP feature on the Contr head. Given that P is not the head of D's complement, it is the (remnant) PP that moves to Spec,D and hence D cannot M-merge with P. D will M-merge instead with one or all the heads in its definiteness valuation chain. This chain includes N and A, both of which are part of the CP which sits at the edge of the NumP phase and are therefore visible for Agree with the [def] feature on D. In a language with a strictly local version of M-Merge,

the definite D will M-Merge with, and get spelled out on, N only. The output of (18) in such a language is P-N$_{def}$-A. In contrast, in a language that allows multiple M-Merge, the definite D will M-merge with both N and A and will be spelled out on both. The output of (19) in such a language is P-N$_{def}$-A$_{def}$.
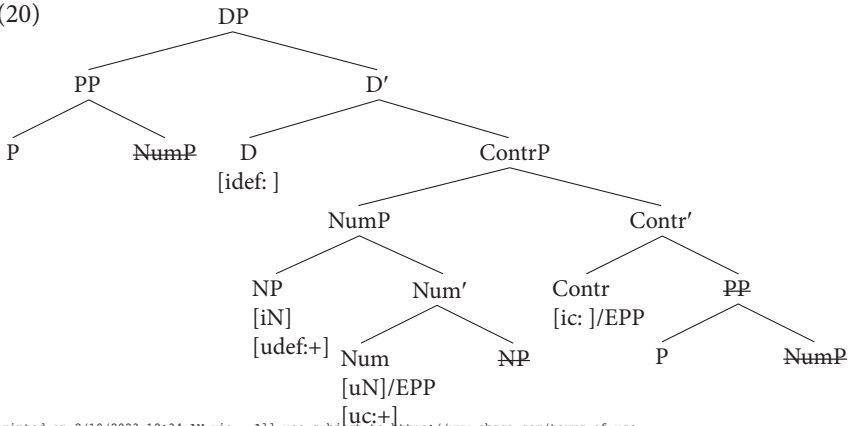
If a prenominal A is merged, the structure is as in (19).

(19)

```
                          DP
            ┌─────────────┴──────────────┐
           PP                            D′
        ┌───┴────┐              ┌─────────┴──────────┐
        P       N̶u̶m̶P̶          D                  ContrP
                            [idef: ]         ┌──────────┴────────────┐
                                          NumP                    Contr′
                              ┌────────────┴─────────┐         ┌────┴─────┐
                             FP                     Num′     Contr       P̶P̶
                        ┌─────┴──────┐          ┌────┴───┐  [ic: ]/EPP ┌──┴───┐
                       AP           F′         Num      F̶P̶            P    N̶u̶m̶P̶
                      [ua: ]   ┌─────┴────┐   [uN]
                      [udef: ] F         NP  [uc:+]/EPP
                              [uc:+]    [ia:+]
                                        [udef:+]
```

The derivation based on this structure is very similar to the one in (18). The only difference is that within the FP that sits in Spec,Num the A precedes N, rather than following it. The output strings of this structure are therefore either P-A$_{def}$-N (in a language with local M-Merge) or P-A$_{def}$-N$_{def}$ (in a language that allows spread).
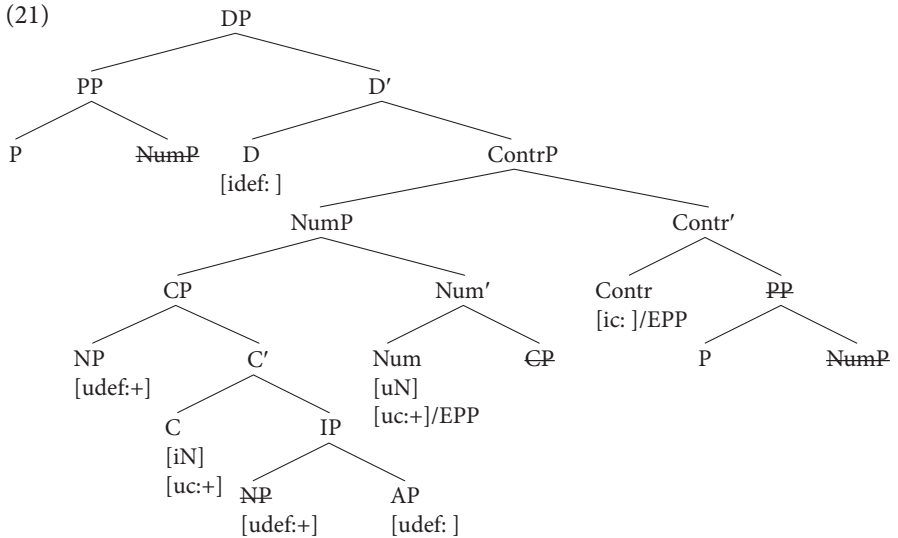
5.1.4   *Num₄: [uN]/EPP,[uc:+]*

Finally, if a Num head with an EPP on the [N] feature is merged, Num will always attract the NP to its Spec, unless a closer item with an [N] feature is present, such as CP or AP. If no A is present, the structure is as in (20).

(20)

```
                          DP
            ┌─────────────┴──────────────┐
           PP                            D′
        ┌───┴────┐              ┌─────────┴──────────┐
        P       N̶u̶m̶P̶          D                  ContrP
                            [idef: ]         ┌──────────┴────────────┐
                                          NumP                    Contr′
                              ┌────────────┴─────────┐         ┌────┴─────┐
                             NP                     Num′     Contr       P̶P̶
                            [iN]              ┌──────┴────┐  [ic: ]/EPP ┌──┴───┐
                            [udef:+]         Num         N̶P̶           P    N̶u̶m̶P̶
                                            [uN]/EPP
                                            [uc:+]
```

In (20) the NP raises to Spec,Num and the whole NumP then moves to Spec,Contr as a consequence of its [c] feature. Given that P cannot move to Spec,D as a head, D will not M-merge with P but with one or all the heads in its definiteness valuation chain. Since D's definiteness valuation chain in this case contains only N, D will M-merge with N and the definite article will be overtly expressed on the N. The only possible output string from a structure like (20) is thus P-N_{def}.
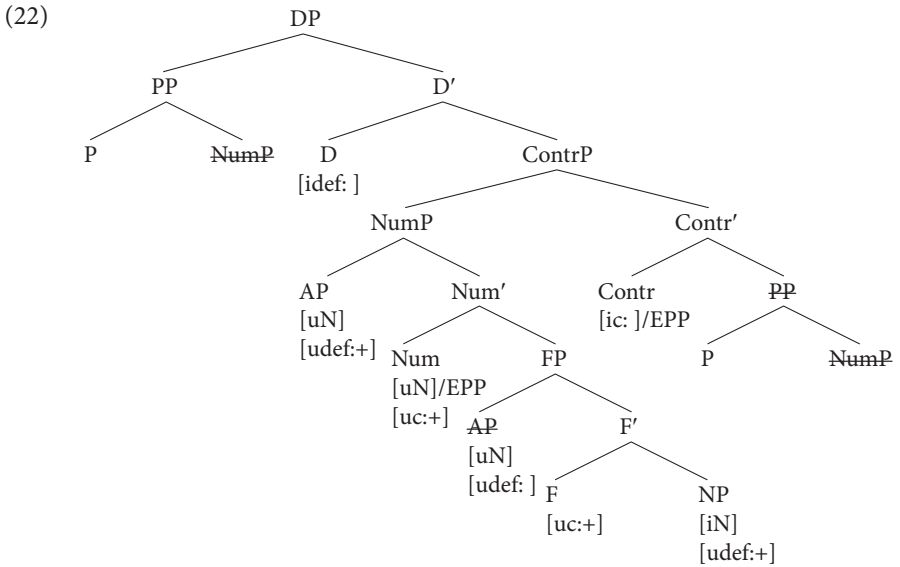
If a postnominal A modifier is present, the structure is as in (21).

(21)

```
                              DP
                  _____
                 PP                        D′
              ____|____             _____|_____
             P        NumP         D               ContrP
                               [idef: ]       _____|_____
                                            NumP              Contr′
                                      _____|_____     _____|_____
                                     CP           Num′   Contr      PP
                                  ___|___      ___|___  [ic: ]/EPP  _|_
                                 NP     C′    Num    CP            P  NumP
                               [udef:+] |   [uN]
                                     ___|___ [uc:+]/EPP
                                    C      IP
                                   [iN]  __|__
                                   [uc:+] NP   AP
                                        [udef:+] [udef: ]
```

The constituent that moves to Spec,Num in (21) is no longer the NP, because a closer item with a matching [N] feature is present, namely CP. Apart from that, the derivation is similar to the one in (20). Given that this time D's definiteness valuation chain contains more than one head, the strings that can result from a structure like (21) are either P-N_{def}-A (for a language in which M-Merge is strictly local) or P-N_{def}-A-$_{def}$ (for a language that allows multiple M-merge).

Finally, if a prenominal A is present, the structure is as in (22).

In (22) the Num head attracts the closest item with an [N] feature, i.e. the AP and subsequently, the NumP raises to Spec,Contr to check its [c] feature. Given that P cannot raise to Spec,D as a head, the definite D will not M-Merge with P but with one or all the heads in its definiteness valuation chain. Since only the AP is attracted to the edge of the NumP phase, only the A will be 'visible' for Agree when the definite D searches for a matching [def] feature. Hence only A is part of the definiteness valuation chain, and the definite D must M-merge with A. The only possible output string of (22) is thus P-A_{def}-N.

(22)

```
                              DP
                 ┌─────────────┴──────────────┐
                PP                            D′
            ┌────┴────┐              ┌─────────┴──────────┐
            P        N̶u̶m̶P̶           D                  ContrP
                              [idef: ]        ┌───────────┴────────────┐
                                           NumP                     Contr′
                                    ┌────────┴────────┐        ┌──────┴──────┐
                                   AP              Num′      Contr          P̶P̶
                                  [uN]         ┌────┴────┐   [ic: ]/EPP   ┌───┴────┐
                                  [udef:+]    Num       FP               P      N̶u̶m̶P̶
                                           [uN]/EPP  ┌───┴───┐
                                           [uc:+]   A̶P̶      F′
                                                   [uN]  ┌──┴───┐
                                                   [udef: ] F     NP
                                                        [uc:+]  [iN]
                                                               [udef:+]
```

To sum up, the various possible feature specifications on the Num head discussed above will generate the following strings illustrated in Table 4:

Table 4. PP strings generated by each type of Num

	Local M-merge Lgs	Spread M-merge Lgs
Num_1		P-N
Num_2		$P\text{-}N_{def}$
		$P\text{-}N_{def}\text{-}A$
		$P\text{-}A_{def}\text{-}N$
Num_3	$P\text{-}N_{def}\text{-}A$	$P\text{-}N_{def}\text{-}A_{def}$ $P\text{-}A_{def}\text{-}N_{def}$
	$P\text{-}A_{def}\text{-}N$	
Num_4	$P\text{-}N_{def}$	$P\text{-}N_{def}$
	$P\text{-}N_{def}\text{-}A$	$P\text{-}N_{def}\text{-}A_{def}$ $P\text{-}A_{def}\text{-}N_{def}$
	$P\text{-}A_{def}\text{-}N$	

In what follows I show how an analysis based on the assumptions above can account for the distribution of the overt definite D in objects of Ps in ER languages. The discussion is structured by languages rather than by the patterns illustrated in Tables 1–3. This is because theoretical decisions about how to account for a pattern in a particular language has consequences on the analysis of the other patterns in that same language.

5.2   Romanian

As shown in Tables 1–3, the strings that can be generated in Romanian are P-N (for unmodified nouns), P-N$_{def}$-A and P-A$_{def}$-N (for Ns with prenominal and postnominal modifiers, respectively).

(23) a. *în parc(-*ul).*   (R)
          in park(-DEF)
          'in the park'

      b. *în centr-ul   vechi*   (R)
          in center-DEF old
          'in the old town'

      c. *Ne-am   plimbat **pe frumoase-le   alei**.*   (R)
          CL.1plAcc-have walked on beautiful-DEF alleys
          'we walked on the beautiful alleys.'

Given that definiteness is never overtly expressed on both Ns and As at the same time in R, I will assume that M-merge in R is strictly local. Moreover, I propose that the Num head in R can bear either an [a] feature or a [c] feature, but not both. In sum, the types of Num heads that define R objects of Ps are Num$_1$ (which yields P-N strings, as explained in § 5.1.1.) and Num$_3$ (which generates P-N$_{def}$-A and P-A$_{def}$-N strings).

5.3   Megleno-Romanian and Istro-Romanian

MR and IR pattern identically with respect to the overt realization of the definite article in objects of Ps, as captured in Tables 1–3. The proposed analysis will therefore apply to both of these languages. The strings that can be generated in MR and IR are P-N, and P-N$_{def}$ (for unmodified Ns), P-N$_{def}$-A$_{(def)}$ and P-A$_{def}$-N$_{(def)}$ (for modified Ns). For lack of space, I have included below only examples from MR.

P-N (MR, Zegrean 2012: 38)
(24) *Ubides   capelā cari   s-mi   veagljā **di   soare***
      loof.for/1s hat   which subj-CL protect.3s from sun
      'I'm looking for a hat that would protect me from the sun'

P-N$_{def}$ (MR, Saramandu et al. 2011: 195)
(25) *Mulárę   si   zācátā și   **cútru   soar-li***
      woman.DEF cl.refl looked also towards sun-DEF
      'The woman looked towards the sun as well.'

P-N$_{def}$-A$_{def}$ (MR, Capidan 1935: 124)
(26) *Și   șa, **la fakir-u   urač-u**   furi-l   al la   furară șimidoil boi.*
      And so, to peasant-DEF poor-DEF thieves-DEF to him stole   both   oxen
      'And so, the thieves stole both oxen from the poor peasant'

P-N$_{def}$-A (MR, Saramandu et al 2011: 188)

   (27)   *cu   fustán-u   mári*
          with skirt-DEF big
          'with the big skirt'

P-A$_{def}$-N$_{def}$ (MR, Saramandu et al 2011: 198)

   (28)   *Feş   doi ań   **la gărţésc-a şcolă**.*
          Did.1sg two years at Greek-DEF school
          'I attended the Greek school for two years'

One important difference between R on the one hand and MR and IR on the other is that definiteness can be overtly expressed on multiple heads in MR and IR. I will therefore assume these languages allow multiple M-merge with all the heads in the definiteness valuation chain.

    The strings attested in MR and IR can be generated under the assumption that the Num head in these languages always bears a periphery feature and moreover that the periphery feature is associated with an EPP. More specifically, given Table 4, the possible feature specifications of Num for these languages are Num$_1$, which will generate P-N strings; Num$_2$, which will generate P-N$_{def}$, P-N$_{def}$-A, and P-A$_{def}$-N strings; and Num$_3$, which will generate P-Ndef-Adef and P-Adef-Ndef in languages such as MR and IR which allow definiteness spread.

5.4   Northern Aromanian (NAr)

NAr is similar to MR and IR, with the exception of P-A$_{def}$-N$_{def}$ strings, which are not attested in NAr.

P-N, P-N$_{def}$ (Tomic 2006: 185)

   (29)   *Bag-u **pi masă/mas-a***
          Put-it on table/table-DEF
          'Put it on the table'

P-N$_{def}$-A (NAr, Saramandu 2007: 393)

   (30)   *nă   dúsim **la consul-u   armănescu***
          CL.1pl.Acc went   to consul-DEF Romanian
          'We went to the Romanian consul'

P-N$_{def}$-A$_{def}$ (Cândroveanu 1977: 190)

   (31)   *fărtat-lu   aţel   mar-le aclimă **pri fărtat-lu   aţel   ñic-lu***
          brother-DEF AŢEL big-DEF called at   brother-DEF AŢEL little-DEF
          'The big brother called the little brother'

P-A$_{def}$-N (Cândroveanu 1977: 358)

   (32)   *paplu Nastu [] avina cu   mutriţa **pri tinir-lu   xen***
          old   Nastu [] eyed with glance at young-DEF stranger
          'Old Nastu furtively glanced at the young stranger'

In order to account for the absence of P-A$_{def}$-N$_{def}$ strings in NAr, I will assume that the Num head in NAr cannot bear the features that yield P-A$_{def}$-N$_{def}$ strings, in other words NAr lacks Num$_3$. Apart from that, NAr has Num$_1$ and Num$_2$, just as MR and IR. Given this choice of Num heads, in NAr the EPP feature is always associated with the [a] feature, and never with the [c] feature on Num, and NAr has Num$_1$ and Num$_2$ only. In order to account for the presence of P-N$_{def}$-A$_{def}$ strings, which would be generated by Num$_3$, I will start from the observation that these strings always include *aţel*, an item that I will analyze as a demonstrative. I follow Cornilescu (2005) and Giusti (2005), among others, in assuming that postnominal demonstratives are merged in a DemP lower than Num. I will also assume that the Dem head always attracts the NP to its Spec, to check a [ua:]/EPP feature on Dem. The structure I propose for P-N$_{def}$-A$_{def}$ strings in NAr is as in (33):

(33)
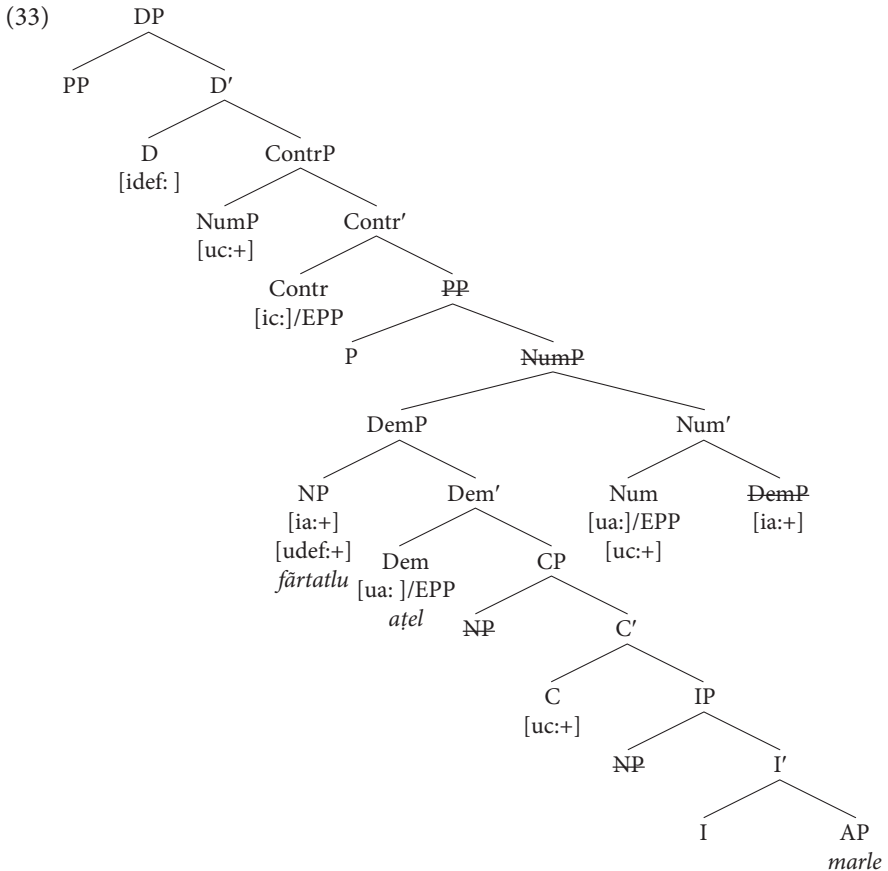
In (33) the postnominal A is merged as usual within a reduced relative clause. The NP is initially merged within this reduced relative clause, as the subject of the AP predicate, and then it raises to Spec,Dem, to check the latter's EPP feature. If Num_1 is merged, which bears an [ua:]/EPP feature only, Num will attract the DemP to its Spec (the closest item with a matching [a] feature), but the derivation will crash because the [uc] feature on C will remain unchecked. On the other hand, if Num_2 is merged, which bears both an [ua:]/EPP feature and a [uc:+] feature, Num again will attract DemP to its Spec and NumP will be attracted to Spec,Contr to check the latter's EPP feature, as represented in (33). The definite article will be spelled out on both N and A in this case, given that NAr allows multiple M-merge. Both N and A are contained in the DemP placed at the edge of Num, and hence they are both 'visible' when D searches for a valued [def].

The reason why *aţel* is obligatory in (P)-N_{def}-A_{def} strings is that in the absence of a DemP, what Num would attract to its Spec would be the NP (the closest item with a matching [a] feature, that could check the EPP on Num). In other words, in the absence of Dem, the derivation in (33) would be equivalent to the one in (16) and the definite article would be spelled out on N only.

5.5   Southern Aromanian (SAr)

Just like MR, IR, and NAr, SAr allows the spread of M-merge to all the heads in the definiteness valuation chain, and thus the definite article can be overtly expressed on multiple heads within the DP. The strings that are attested in SAr are P-N, P-N_{def}, and P-N_{def}-A_{def}.

P-N (Cândroveanu 1977: 212)
(34) *păduri-le   ligănate **di vimtu***
      forests-DEF swayed by wind
      'the forests swaying in the wind'

P-N_{def} (Cândroveanu 1977: 232)
(35) *haide   **la capidan-lu***
      let's.go to captain-DEF
      'Let's go to the captain'

P-N_{def}-A_{def} (Campos 2005: 313)
(36) ***puľi-ľi   (atseľi)   ňits-ľi***   li   adună   mîsa
      birds-DEF those   small-DEF them gathered mother
      'The mother gathered the small birds.'

In order to account for the SAr facts, I propose that (i) the types of Num heads that define SAr objects of Ps are Num_1 (which yields P-N strings) and Num_4 (which generates $P-N_{def}$, $P-N_{def}-A_{def}$ and $P-A_{def}-N_{def}$ strings), and (ii) As in SAr are never merged in a left periphery position and thus never precede Ns. Therefore, even though from the point of view of the features present on Num_4 $P-A_{def}-N_{def}$ strings should be possible, these strings are ruled out independently and as a result only $P-N_{def}$ and $P-N_{def}-A_{def}$ strings are generated by Num_4.

5.6 To sum up

The discussion in § 5 showed that the overt expression of the definite article in ER objects of Ps depends on two factors: (i) whether the respective language has a strictly local version of M-merge, in which only the closest head to D that bears a [def] feature can bear the overt definite article, or multiple M-merge, in which case the definite article can be expressed on all the heads that bear a [def] feature within the DP; and (ii) the exact feature specification of Num in the respective language. With respect to the first factor, Romanian is the only ER language in which the definite article is spelled out in a strictly local manner, on the head that is closest to D in the definiteness valuation chain (i.e. the closest item with a [def] feature), while all the other ER languages display 'definiteness spread', i.e. the definite article is spelled out on each and every head in the definiteness valuation chain. With respect to the second factor, ER languages differ as in Table 5:

Table 5. Types of Num in ER languages

		R	MR/IR	NAr	SAr
Num_1	[uN] [ua:]/EPP	✓	✓	✓	✓
Num_2	[uN] [ua:]/EPP [uc:+]		✓	✓	
Num_3	[uN] [uc:+]/EPP	✓	✓		
Num_4	[uN]/EPP [uc:+]				✓

Table 5 shows that in all ER languages the Num head can have an [a] feature (all languages have Num_1), and that whenever the [a] feature is present, it is associated with an EPP. In contrast, only in some of the ER languages does the Num head have a [c] feature (either as the only periphery feature or in association with the [a] feature). Moreover, the [c] feature may, or may not, be associated with an EPP.

6.   Conclusions

This chapter has investigated the properties of definite objects of Ps in Eastern Romance languages. Objects of Ps in these languages display a peculiar pattern: the definite article can be null if objects of Ps are unmodified, but with modified objects of P, the definite article must be expressed overtly (either on the N, or on the A modifier, or on both). In spite of these shared properties, ER languages display microvariation with respect to which particular item in the DP can host the definite article, and with respect to whether the definite article in unmodified objects of Ps must, or simply can, be null.

The chapter argues that it is not necessary to enrich the typology of determiners in Romance so as to allow for null definite Ds. Rather, we argued that null definite Ds and overt definite Ds are identical from a morpho-syntactic point of view and that the overtness of the D is an outcome of a Spell-out rule. According to the proposed rule the definite D is spelled out overtly only if D M-merges with ahead that bears a [def] feature. The covertness of the definite article in unmodified objects of Ps in ER was accounted for by showing that in these cases D M-merges with P, which does not bear a [def] feature. The observed microvariation across ER was accounted for by two factors: (i) the feature content of the Num head, in particular whether the Num head bears periphery features like [a] and [c] or not, and whether an EPP feature is associated with these periphery features on Num; and (ii) the domain of application of the M-merge and Spell-out rules we proposed.

References

Alexiadou, Artemis. 2001. *Functional structure in nominals: Nominalization and ergativity*. Linguistik Aktuell/Linguistics today 42. Amsterdam: John Benjamins. https://doi.org/10.1075/la.42

Campos, Hector. 2005. Noun modification, pseudo-articles, and last resort operations in Arvantovlaxika and in Romanian. *Lingua* 115: 311–347. https://doi.org/10.1016/j.lingua.2003.11.002

Cândroveanu, Hristu. 1977. *Antologie de proză aromână*. Bucharest: Editura Univers.

Capidan, Theodor. 1935. *Meglenoromânii – Dicționar meglenoromân*, Volume 3. Bucharest: Cultura Națională/ Academia Română. Studii şi Cercetări XXV.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In David Michaels Martin & Juan Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A life in language*, 1–52. Cambridge, Mass: MIT Press.

Cinque, Gugliemo. 2010. *The Syntax of Adjectives*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262014168.001.0001

Cornilescu, Alexandra. 2004. Relative Adjectives. *Bucharest Working Papers in Linguistics* 6 (1): 50–73.

Cornilescu, Alexandra. 2005. Romanian demonstratives and minimality. *Bucharest Working Papers in Linguistics* 7(1): 102–116.

Cornilescu, Alexandra & Alexandru Nicolae. 2009. On the syntax of Romanian definite phrases. Changes in the patterns of definiteness checking. In Petra Sleeman & Harry Perridon (eds.), *The Noun Phrase in Romance and Germanic: Structure, variation, and change*, 193–221. Amsterdam: John Benjamins. https://doi.org/10.1075/la.171.14cor

Cornilescu, Alexandra & Alexandru Nicolae. 2011. Nominal Peripheries and phase structure in the Romanian DP. *Revue Roumaine de Linguistique* 56(1): 35–68.

Cornilescu, Alexandra & Anca Dinu. 2013. Adjectives and specificity. *Revue roumaine de linguistique* 58(4): 455–480.

Dobrovie-Sorin, Carmen & Ion Giurgea. 2006. The Suffixation of Definite Articles in Balkan Languages. *Revue roumaine de linguistique* 51. 73–104.

Giusti, Giuliana. 2005. At the left periphery of the Romanian noun phrase. In Martine Coene & Liliane Tasmowski (eds), *On space and time in language*, 23–49. Cluj-Napoca: Clusium.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Blackwell: Malden.

Higginbotham, James. 1985. On Semantics. *Linguistic Inquiry* 16, 547–593.

Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass: MIT Press

Lopez, Luis. 2009. *A Derivational Syntax for Information Structure*. Oxford, OUP. https://doi.org/10.1093/acprof:oso/9780199557400.001.0001

Mardale, Alexandru. 2006. Why *on table* is *on the table*? Notes on the definite article's incorporation with prepositions in Romanian. In Beata Gyuri (ed.), *Proceedings of the First Central Europe Student Conference in Linguistics*, 1–12. Budapest: Pázmány Péter Catholic University.

Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37:69–109. https://doi.org/10.1162/002438906775321184

Pesetsky, David & Esther Torrego. 2001. T-to-C Movement: Causes and Consequences. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language*, 355–426. Cambridge, MA: MIT Press.

Pesetsky, David & Esther Torrego. 2004. Tense, Case and the Nature of Syntactic Categories. In Jacqueline Guéron and Jacqueline Lecarme, and Alexander Lecarme (eds.), *The Syntax of Time*, 495–538. Cambridge MA: MIT Press.

Pesetsky, David & E. Torrego. 2007. The Syntax of Valuation and the Interpretability of Features. In Simin Karimi, Vida Samiian and Wendy K. Wilkins (eds.), *Phrasal and Clausal Architecture: Syntactic Derivation and Interpretation. Linguistik Aktuell/Linguistics Today* 101, 262–294. Amsterdam: John Benjamins. https://doi.org/10.1075/la.101.14pes

Saramandu, Nicolae. 2007. *Aromâna vorbită în Dobrogea*. Bucharest: Editura academiei române.

Saramandu, Nicolae, Marilena Tiugan, Irina Floarea & Alina Celac. 2011. Dicționar meglenoro-mân. *Fonetică și dialectologie*, 30. 51–115. Bucharest: Editura Academiei.

Tănase-Dogaru, Mihaela. 2012. *The Syntax of Quantity and Quality in Romanian*. Bucharest: Editura Universității.

Teodorescu, Alexandra. 2006. Adjective ordering restrictions revisited. In Donald Baumer, David Montero & Michael Scanlon (eds.), *Proceedings of the West Coast Conference on Formal Linguistics* (WCCFL 25), 399–407. Somerville, MA: Cascadilla Press.

Tomic, Olga. 2006. *Balkan Sprachbund Morphosyntactic Features*. Dordrecht: Springer. https://doi.org/10.1007/1-4020-4488-7

Zegrean, Iulia-Georgiana. 2012. Balkan Romance: Aspects on the Syntax of Istro-Romanian. Venice, Italy: Università Ca'Foscari Venezia dissertation.

Bridging issues at the CP-TP-vP levels

Differential object marking, oblique morphology, and enriched case hierarchies

Monica Alexandrina Irimia

University of Modena and Reggio Emilia

The puzzle of oblique morphology on differentially marked objects (DOM) has received renewed attention in the recent formal literature, under two main theoretical lines: (i) oblique syntax for DOM (Manzini & Franco 2016, 2019, inter alii); (ii) oblique marking on DOM as morphological syncretism (Keine & Müller 2008; Keine 2010; Bárány 2018, *inter alii*). This paper evaluates the predictions made by these accounts against a limited set of Romance DOM varieties, crucially including both DOM as dative (Western Romance) and DOM as a non-dative adposition (Romanian). A specific syntactic implementation for oblique DOM is proposed, building on the idea of additional licensing in certain types of accusatives, combined with enriched case hierarchies (Starke 2017). This not only preserves the advantages of the above-mentioned accounts, but also opens up the path to addressing several less-studied or recalcitrant (micro)variation points.

Keywords: differential object marking, oblique morphology, syncretism, enriched case hierarchies

Introduction

Under the broad spectrum of differential object marking (DOM), referential animate direct objects require a preposition across Romance. In Western varieties, this morphology is syncretic with the dative (Jaeggli 1982; Bossong 1991; López 2012; Manzini & Franco 2016, a.o.), as seen in the Standard Spanish example in (1a). Romanian DOM also uses a preposition, namely the locative *pe* "on" in (1b) (Dobrovie-Sorin 1994; Cornilescu 2000; Mardale 2008; Tigău 2011, a.o.).

(1) a. *He   encontrado* *(*a*)   *la   niña*/(*a)   *el   libro.*
        have.1SG found   DAT=DOM DEF.F.SG girl/DAT=DOM DEF.M.SG book
        "I have found the girl/the book."

                                 (Spanish, Ormazabal & Romero 2013, 1a, b)

b. *(I)-am*　　　　　*văzut (pe)*　　*băieţii*　　*înalţi*　/(**pe*)
CL.3M.PL.ACC-AUX.1 seen LOC=DOM boy.DEF.M.PL tall.M.PL /DOM
pereţii　　*verzi.*
wall.DEF.PL green.PL
"I have seen the tall boys/the green walls."　　　　　(Romanian)

Signaling the most typical 'structural' objects via 'non-canonical', 'oblique' marking is a non-trivial puzzle going beyond Romance (Bossong 1991; Irimia & Pineda 2019, a.o.), and has received renewed attention in the recent formal literature. We examine here two prominent lines of inquiry, namely (i) oblique syntax for DOM (Manzini & Franco 2016, 2019, a.o.) and (ii) morphological resolutions of the OBL-DOM syncretism (Keine & Müller 2008; Bárány 2018, a.o.). The emerging conclusion is that Romance DOM and its microvariation cannot be fully captured under either of these two implementations.

The solution proposed here builds on theories equating oblique DOM to an 'additional licensing operation' on certain types of accusatives. We see this additional licensing operation as connected to a discourse-anchoring mechanism, which includes the syntactic encoding of perspective/sentience (Pancheva & Zubizarreta 2018, a.o.). Oblique morphology signals the presence of this additional licenser in the same local licensing domain as an initial licenser for accusatives. This predicts co-occurrence of DOM with other structural licensing mechanisms for direct objects as well as the pervasive accusative syntactic behavior of such objects; additionally, via the use of enriched case hierarchies (Starke 2017), it can also derive the distinct syncretism patterns DOM and various (structural) accusatives establish across Romance.

The structure of the paper is as follows. Section 1 addresses the DOM-DAT unification as syntactic obliques, and Section 2 applications of the morphological accounts to the various types of Romance accusatives. The two sections also present some problems raised by each of these formalizations. Section 3 introduces the additional licensing hypothesis, as well as the advantages of enriched case hierarchies (building on Starke 2017 and Caha 2009). Section 4 contains the conclusions.

1.   Differential objects as syntactic obliques

A recent proposal by Manzini & Franco (2016) is that oblique morphology on DOM is not a mere instance of surface opacity; it, in fact, signals DOM membership into a larger syntactic category which introduces a part-whole relation, and which is also characteristic of obliques.

The authors see the oblique case as an elementary predicate/operator, semantically specified for possession/inclusion and notated as Q $_{(\subseteq)}$, as in (2a). Turning now to DOM (e.g., the Romanian sentence in (1b)), the explicit proposal Manzini & Franco (2016) make is that internal arguments specified with Participant features also imply the presence of a QP $_{(\subseteq)}$ constituent, as in the adapted representation in (2b); this element is required by the referential properties of this type of internal argument.
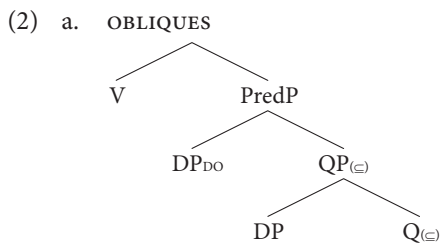
(2)   a.   OBLIQUES

    b.   Romanian DOM

           (Manzini & Franco 2016: Example 24/31, p. 213/218, adapted)
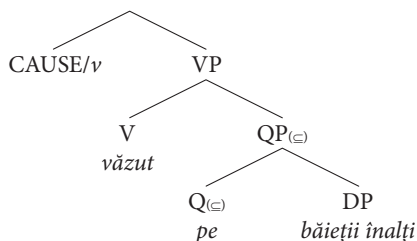
Due to the QP $_{(\subseteq)}$ structural similarity, DOM is assumed to always behave like true obliques syntactically. The challenge, however, is that syntactic diagnostics group DOM (the traditional 'prepositional accusative') with ACC, excluding DAT/OBL. Table 1 summarizes some of these tests.[1]

Manzini & Franco (2016) are, of course, aware of the numerous syntactic diagnostics severing DOM from DAT/OBL, but the working hypothesis is to filter out them independently. However, some of the remarks they make in this direction prove orthogonal to the point and appear not to resolve the ACC-DOM uniformity.

---

[1].   The language sample has been selected so as to allow a basic coverage of the multiple facets of DOM in its interaction with nominal/sentential syntax. Not all the examples can be included in this very short paper. Elicitations from native speakers do not have a source cited. Other syntactic diagnostics (distinguishing DOM from OBL) are discussed throughout the paper or can be found in the literature (see especially Bárány 2018 for a comprehensive list).

Table 1. ACC, DOM and OBL syntactic diagnostics

Diagnostic & language	ACC	DOM	DAT/OBL
PPA (Past Participle Agreement) (Neapolitan, (Old) Catalan (pronouns), Sardinian)	✓	✓	*
(Periphrastic) Passivization (Catalan, Old Catalan, Neapolitan, Romanian, Standard Spanish, Sardinian, leísta Spanish)	✓	✓	*
ACC *clitic doubling of direct object pronouns* (Standard Spanish, Catalan, Old Catalan, Romanian)		✓	*
ACC *clitic doubling of direct object DPs* (Neapolitan)		✓	*
Case preserved under nominalization (Catalan, Neapolitan, Romanian, Sardinian, Standard Spanish, leísta Spanish)	*	*	✓
Hosting secondary predicates (Catalan, Neapolitan, Romanian, Sardinian, Standard Spanish, leísta Spanish)	✓	✓	* (or very restricted)
Hosting reduced relative clauses Catalan, Neapolitan, Romanian, Sardinian, Standard Spanish, leísta Spanish)	✓	✓	*

A prominent syntactic test canonically used to set aside DOM from OBL is their divergent behavior under periphrastic passivization (PASS). As shown in Table 1, regular ACC and DOM pattern alike in all the varieties examined here, to the exclusion of the goal DAT/OBL. The former two allow passivization, with change to NOM and promotion to subject position, while the goal DAT and obliques fail this operation. Manzini & Franco (2016) put forward the conclusion that periphrastic passivization does not undermine the syntactic unity of DOM and OBL as it is due to independent conditions affecting accusatives. These conditions are, however, not discussed.

One piece of evidence Manzini & Franco (2016) do address comes from instances of so-called 'extended accusatives', i.e., accusative morphology which is preserved under passivization (with possible promotion from object to subject). An example comes from standard Spanish (3), where, as opposed to the periphrastic PASS, impersonal SE (SE_IMP) can preserve not only the DAT, but also DOM (and ACC).

(3) a.   *Se   me   responde.*
          SE_IMP I.DAT reply.3SG
          "They reply to me."

    b.   *Se   me   busca//   Se busca   a   las niñas.*
          SE_IMP I.ACC search.3SG// SE search.3SG DAT=DOM DEF girls
          "They are looking for me/the girls."   (Spanish)

For Manzini & Franco (2016), the SE test indicates the following: (i) (EPP) fronting is not always dependent on structural ACC Case suppression; (ii) the (displaced) DP can preserve its licensed Case (structural or oblique). However, this last property is what makes the test irrelevant to the present discussion, as it does not indicate whether ACC, DOM and DAT are distinct or the same syntactic category. And, in fact, when structural constraints are imposed on SE, such as with the Romanian medio-passive SE$_{MP}$ (Dobrovie-Sorin 1998, a.o.), DOM diverges from DAT (and all obliques). The latter can be promoted to subject position and preserve their oblique morphology as in (4a/b), while DOM must be suppressed, as in (4c). A nominative subject is possible ((4d), as detected by plural agreement):

(4) a. *Ni   se   răspunde.*
        CL.1PL.DAT SE$_{MP}$ răspunde.3SG
        "They reply to us."

    b. *Se   scrie   pe   tablă.*
        SE$_{MP}$ write.3SG LOC blackboard.
        "They write on the blackboard."

    c. **Se   înscrie/înscriu **pe**   elevi.*
        SE$_{MP}$ register.3SG/PL LOC=DOM students
        Intended: "The students get registered."

    d. *Elevii   se   înscriu.*
        students.DEF$_{NOM}$ SE$_{MP}$ register.3PL
        "The students get registered."   (Romanian)

Another troublesome diagnostic for the DOM–OBL common syntax is object agreement. Recent research has confirmed the robustness of overt agreement with (in-situ) direct objects across Romance (Manzini & Savoia 2005; Loporcaro 2010; D'Alessandro & Roberts 2010; Belletti 2017, a.o.). In Example (5) from Neapolitan (adapted from Loporcaro 2010), the (phi-)features of the direct object are tracked by past participle agreement (PPA).[2]

(5) (L')   *addʒə* *kwottə/√ *kɔttə   a*[3]   *pasta* / *a*
        CL.3ACC have.1 cooked.M.SG/F.SG DEF.F.SG pasta / DAT=DOM
        l'aragostə.
        DEF.S.SG-lobster
        "I have cooked the pasta/the lobster."   Neapolitan
                              (Loporcaro 2010, adapted; Roberto Petrosino, p.c.)

---

2. As Loporcaro (2010) has convincingly demonstrated, these examples do *not* involve (right) dislocation.

3. In Neapolitan, DOM is homophonous with the (SG) feminine definite article when nominal roots start in a consonant.

Crucially for us, sentence (5) also shows that DOM is unified with ACC in that both surface with PPA, which always excludes goal DAT (Ledgeway 2000; Loporcaro 2010, a.o.).[4] These facts cannot be derived under a systematic oblique syntactic behavior for DOM; to account for them, Manzini & Franco (2019), instead, explore a labeling explanation. More precisely, although oblique DOM contains a Q $_{(\subseteq)}$ head, the latter does not project and the nominal is labeled as a DP. This assumption could provide us with a way out; but the problem is that there are many instances, such as DOM blocking of overt definiteness, as seen in Romanian (Dobrovie-Sorin 1994) or Sardinian (Jones 1993), where DOM feature labelling and projection both appear to be necessary. A way to reconcile this non-uniform labeling behavior of the Q $_{(\subseteq)}$ head is not easily available.

To conclude, the DOM-OBL syntactic unification does not hold attentive scrutiny, at least under Manzini & Franco's (2016) or (2019) model. There is also the question why the QP $_{(\subseteq)}$ component must imply a *global* oblique behavior for DOM, as opposed to just *some* oblique characteristics. Nevertheless, one intuition proves very useful and has to be maintained, namely that the *complex structural* make-up of differential objects has important syntactic consequences (besides the obvious morphological ones).

2.   Oblique morphology as morphological syncretism

Given the undeniable syntactic differences between DOM and obliques and the difficulties with independently explaining them under a common DOM-OBL syntax, a second, recent, theoretical model, the morphological one, attributes the DAT-OBL homophony to morphological syncretism.

Accounts in this direction include Keine & Müller (2008); Keine (2010), or Bárány (2018), a.o. Despite differences in technical implementation, the general theoretical background is that cases are not atomic entities, but rather decompose into more primitive, hierarchically organized features (Caha 2009; Harðarson 2016; Starke 2017, a.o); syncretism boils down to underspecification in a model such as Distributed Morphology (see also Keine 2010 for DOM in DM, a.o.).

We briefly present two accounts. In an analysis addressing the DOM-DAT homomorphism, Bárány (2018) assumes the hierarchy of case features in (6), where ACC and DAT are contiguous, but DAT is more specified than ACC, as in (7); it shares some features with ACC, but also has distinct features (represented here with highly abstract labels like A, B, C):

---

4.   These patterns also hold diachronically across Romance.

(6)   NOM > ACC > DAT > GEN > LOC > ABL/INS > ….   (Harðarson 2016)

(7)   Case features: ACC = [A, B] DAT = [A, B, C]   (Bárány 2018: example 42, p. 19)

A possibility is for each case marker to be spelled-out by distinct rules, as in (8a), resulting in distinct case morphology at PF. Syncretism implies that both cases are spelled-out by the 'same' spell-out rule, which renders the structural differences between them opaque on the surface. Crucially, this rule does not prevent individual cases from having a 'distinct syntactic' behavior.

(8)   a.   Spell-out rules for distinct case markers
            [A, B] ↔ /-w/   [A, B, C] ↔ /-x/   (Bárány 2018: example 43/44, p. 20)
      b.   Syncretic spell-out rule
            [A, B] ↔ /-y/

Bárány (2018) applies this reasoning to standard Spanish, where, as we have shown in (1a), referential (definite) animates show DAT morphology, while inanimate DPs use a zero-coded form, which is in fact homophonous with the nominative. Bárány (2018) follows López (2012) in the assumption that DOM is assigned accusative case (abstractly labeled [A, B]) when the object raises to a position above VP, namely to the specifier of α, as in (9a). Non-DOM arguments are left caseless, as they (pseudo-)incorporate into V. IOs are assigned dative case (abstractly labeled [A, B, C]) by Appl. As ACC and DAT are the only internal argument categories that carry case features, a single spell-out rule applies to both, as illustrated in (9b).

(9)

(a)

(b)   Spell-out rules for Spanish
      a. [A, B] ↔ a
      b. [A] ↔ - Ø
      (Bárány 2018: ex. 47, p. 22)

(Bárány 2018: ex. 45, p. 21)
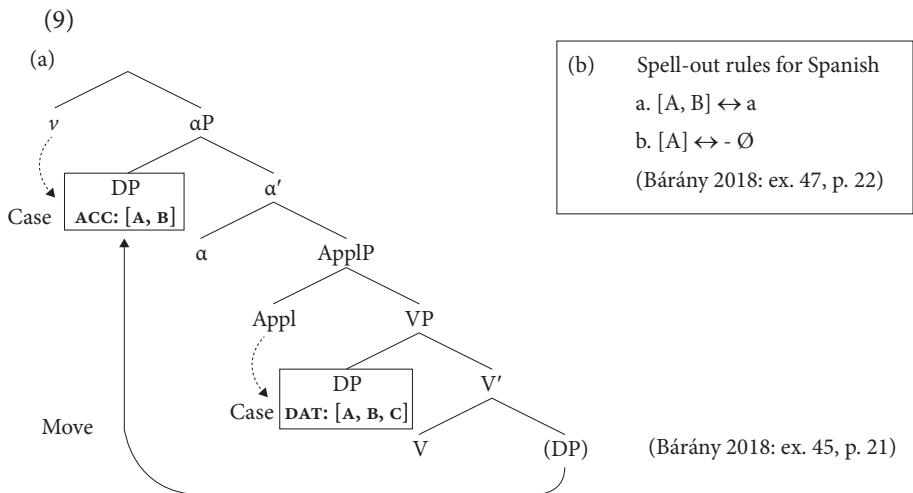
Keine (2010) and Keine & Müller (2008) have a slightly different morphological take on the DOM-DAT syncretism. They see both non-differentially marked objects and DOM as structural accusatives, having been assigned Case. The surface difference results from DOM case morphology being spelled-out, while an impoverishment rule deletes the accusative case features on the non-DOM objects. Thus, the latter are predicted to be spelled out as nominative.

Let us see now what these two systems predict for Romance DOM (beyond Spanish). We can first look at Romanian, where no DOM-DAT homophony obtains; DOM uses a locative preposition, as seen in (1b). Given that syncretism must target 'strictly contiguous' features ('the *ABA Constraint', Caha 2009; Starke 2017, a.o.), Romanian would need a hierarchy similar to (10); a unique spell-out rule will affect the ACC > LOC portion. But this cannot explain the ACC – DAT homomorphism in other categories with case, such as the clitics (which are, in fact, needed to double certain types of DOM); e.g, one exponent of 3 DAT SG CL, namely -i is syncretic with 3 ACC PL, as in (1b), pointing instead to the hierarchy in (6).

(10)   NOM > DAT > GEN > **ACC** > **LOC** > ABL/INS >[5] …. (or …> GEN > **LOC** > **ACC** >…)

In fact, Romanian (accusative) clitics raise another point about Romance, namely the problem of multidimensional argument encoding. More precisely, in the domain of accusatives, it is not the case that the splits are just zero/non-zero. Alternations between two or more 'overt' markers are common, as seen with pronouns. Let us also illustrate with Neapolitan. As we have shown in (5), its direct objects have a more complex morphological behavior than the Spanish ones. Certain types of objects trigger PPA, and animates show DOM and PPA. The syncretism rules discussed in Bárány (2018) take non-DOM direct objects to lack case features as they have not been assigned structural case. This implies that PPA does not result from a licensing operation, begging the question about its presence with DOM. Keine (2010), on the other hand, discusses the problem of PPA, but for languages of the Hindi type; there, PPA is taken to result from a radical impoverishment rule in the syntax. Hindi is, however, crucially different from Neapolitan-type languages in that its DOM blocks PPA. Neapolitan shows both PPA and DOM. Connecting PPA to a radical impoverishment rule (that deletes the accusative case across the board) is not always easy to implement with DOM, so as to avoid the latter ending up both having and lacking ACC.

The assumption that *all* (inanimate) accusatives without DOM are caseless is problematic in yet another respect. Still staying with Neapolitan, we see below

---

5. Also note that GEN and DAT are not fully homophonous in Romanian; the former contains a linker that the latter cannot exhibit (Dobrovie-Sorin 1994, a.o). Thus, both GEN and DAT categories are needed, with GEN lower than DAT.

inanimates that show (obligatory) clitic doubling[6] even in the absence of DOM. Crucially, the clitic part must surface with accusative morphology, and not DOM or DAT. Moreover, this clitic-doubled DP is not only case marked, but also passes 'accusativity' diagnostics (undergoes passivization, etc.).

(11) *L'*   *addʒə* *kwottə/ √ *kɔttə*   *a*   *pastə*.
      CL.3ACC AUX.1 cooked.M.SG/F.SG DEF.F.SG pasta
      "I have cooked the (specific) pasta."   (Neapolitan)

These Neapolitan (and Romanian) data overtly confirm a common Romance pattern – usually, the accusative can be tracked by three exponents: PPA (or Ø),[7] ACC (for pronouns or in ACC clitic-doubling), and DOM. Thus, there can be at least 4 (internal structural) cases in the language, spelled out in different ways: PPA, ACC, DOM, OBL. Given that syncretism affects the case system, the question is under what types of spell-out rules it operates. We can try to annotate the accusative case matrices, using abstract features, such as in (12); we can then add the assumption that morphological rules do not alter/delete full feature bundles, but have access to independent features (a possibility that will avoid the PPA-DOM contradiction):[8]

(12)   PPA = [A, B], ACC = [A, B, C], DOM = [A, B, D], DAT = [A, B, E]

Now the question is: what dictates the DOM – DAT homomorphism? What prevents an output of the type [*DOM DEF pasta have cooked* or *have cooked DOM DEF pasta* …]) for (11), and also the illicit ACC-DAT homomorphism for clitic-doubled DPs (systematically replacing ACC CL with DAT CL in (11)?[9] Note that *some* exponents of the ACC clitic in (11) are actually homophonous with the dative. The question is what blocks 'total homomorphism' with DAT similarly to what is seen in

---

6.   Object licensing by clitic doubling spells-out yet another structural (accusative) case strategy, seen in many languages, even if they do not have adpositional DOM (Anagnostopoulou 2007, a.o.). Neapolitan has been selected here as it is among the languages that have both these licensing strategies (clitic doubling and DOM).

7.   Simplifying the facts here, for reasons of space. There are varieties which exhibit both PPA, as well as Ø-marked accusatives, besides DOM and DAT. This supports the idea that accusatives involve a variety of licensing strategies.

8.   Keine & Müller (2008) analyze a similar system, namely Finnish, where the accusative exhibits four exponents. However, the morphological resolution rules are simpler in Finnish, as the partitive does not have the same status as structural accusatives. The Romance picture is more complex, as at least three exponents have structural case.

9.   The problem is not the restriction of DOM to animacy (see fn. 15 and López 2012 for remarks, a.o.). The issue is not (left-)dislocation either; this process might be possible even with DOM inanimates, but might require a resumptive clitic ([DOM DP, CL$_{ACC}$ [TP…]]).

the ACC-DOM syncretism, where the paradigm is globally affected. In conclusion, in both Romanian and Neapolitan we see that the homomorphism with the dative/oblique can be established in 'two distinct directions': ACC-DAT (clitics, pronouns) and DOM-DAT/OBL. Telling these two options apart can provide insightful hints into the nature of oblique marking for Romance DOM.

3.   Oblique DOM in enriched case hierarchies

To recapitulate, the empirical and theoretical observations presented in the previous two sections point to three main conclusions we need to account for: (i) there are at least three exponents tracking the accusative with overt morphology, namely in PPA, DOM and the ACC clitic part of certain DPs (the problem of multiple overt, structurally licensed accusatives); (ii) DOM is systematically homophonous with exponents of OBL; (iii) these facts cannot be captured under theories which derive the DOM-OBL syncretism based on their being the only categories bearing morphological case features nor under theories which assume an oblique syntax for DOM.

We show here that one promising avenue comes from Starke (2017) who motivates the hierarchy in (13), with two ACC case features, one above DAT/GEN and the other below DAT/GEN. These two accusatives have different structures, which is desirable for the data at hand. However, we also need to encode the locative in our hierarchy, given that Romanian DOM uses precisely this morphology. We can combine Starke (2017) with Caha's (2009) case sequence which does contain the locative, as seen in (14). What we obtain is a so-called 'Enriched Case Hierarchy' as in (15), which can derive not only the DOM-LOC, but also the DOM-DAT, as well as the ACC-DAT syncretism. Given that DOM has a similar profile (sensitivity to animacy, adpositional nature, an anti-incorporation mechanism, etc.) in both Western Romance and Romanian, a single explanation encompassing both types is a welcome result.

(13)   NOM > ACC_1 > DAT_1 > GEN > ACC_2 > DAT_2 > ABL/INS > …

(Starke 2017: example 22; p. 5)

(14)   NOM > LOC_1 > GEN > LOC_2 > DAT > LOC_3 > ABL/INS > …   (Caha 2009: 10, 130)

(15)   NOM > ACC_1 > LOC_1 > DAT_1 > GEN > LOC_2 > ACC_2 > DAT_2 > LOC_2

Enriched Case Hierarchy

3.1   Nominal structure in enriched case hierarchies

One important observation we need to start from is that nominals can be featurally complex in Romance languages; more specifically, they might contain additional features beyond uninterpretable Cae [uC], whose checking normally results in an unmarked form, homophonous with the nominative. One such extra feature is what we abbreviate here as SENTIENCE, which grammaticalizes animacy (or humanness), and which is spelled out via the oblique DOM [10,11] Yet another specification is connected with PERSON$_{CL}$ linked with accusative clitic doubling.

(16) DP$_{DO}$
$$\begin{bmatrix} [_uC] \rightarrow \text{ACC}_1 \\ \text{SENTIENCE} \rightarrow \text{ACC}_2 \\ \text{PERSON}_{CL} \rightarrow \text{ACC}_3 \end{bmatrix}$$

SENTIENCE appears to require licensing below vP. As has been observed by López (2012), DOM cannot bind into the EA (observation confirmed in the languages examined here).[12] Given that the initial licenser in the relevant domain (either v or a low Asp head, see Martín 2005) is used for checking [$_u$C], an additional licenser must be used, as a last resort (adapting Jaeggli 1982,[13] or more recently Kalin 2018). Following both López (2012), as well as Pancheva & Zubizarreta (2018), we can make use of an α projection, located above VP, but below vP. For López (2012), α contains a conglomerate of applicative and aspectual specifications, while for Pancheva & Zubizarreta (2018) it is an Appl head which collapses features related to empathy, sentience, perspectivization and viewpoint. As Romanian DOM is not homophonous with the dative, we can assume that α does not contain applicative features in this language, but only features related to sentience and perspectivization. In fact, given that in Western Romance the DOM *a* is also homophonous with

---

10. Across languages SENTIENCE can be encoded as a PERSON feature, possibly connected with semantic gender (Cornilescu 2000; Rodríguez-Mondoñedo 2007, a.o.). However, an extensive discussion of the typology of PERSON features across Romance and their interaction with DOM is beyond our immediate concern.

11. Given these remarks, a question would be why the marker does not also extend to subjects. A definitive answer to this problem requires, first of all, an investigation of the contexts where the differential marker can be seen with subjects; this issue is beyond the scope of this paper.

12. Note that this does not entail that SENTIENCE (DOM) cannot take wide scope outside vP. As López (2012) discusses, one technical possibility is to associate the differential marker with *Choice Functions*.

13. However, the recruitment of a last-resort licenser is not triggered by competition with the licensing needs of an accusative clitic double.

a locative, a valid question is whether DOM systematically uses, across Romance, a 'locative' strategy (which is otherwise common cross-linguistically). In many Romance varieties, PERSON$_{CL}$ spells out a feature which requires scrambling above v,[14] either to a Case projection (Săvescu Ciucivara 2009, a.o.), or to a PERSON locus (Belletti 2005, a.o.) in the vP periphery. A schematic representation is offered in (17).

(17)

(b) φ $_{u}$C [ACC$_1$] $\leftrightarrow$ PPA/-Ø
       SENTIENCE [ACC$_2$] $\leftrightarrow$ DOM
       PERSON$_{CL}$ [ACC$_3$] $\leftrightarrow$ CLTIC$_{ACC}$

DP$_{DO}$
φ [$_u$C] $\rightarrow$ ACC$_1$
SENTIENCE $\rightarrow$ ACC$_2$
PERSON$_{CL}$ $\rightarrow$ ACC$_3$
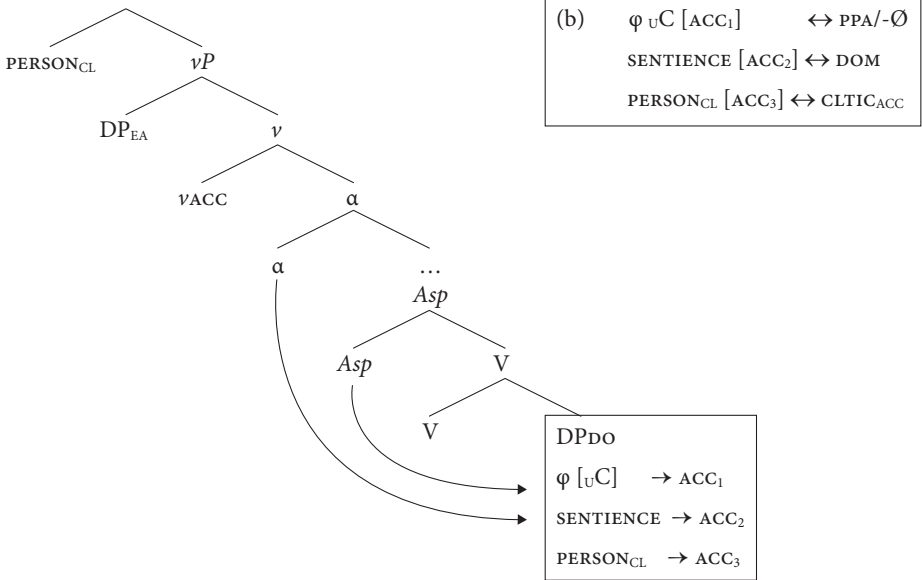
One posssibility would be to assume that the additional licenser is a locative (oblique), and maintain Manzini & Francos's (2016) intuition that animacy is decomposed as a locative/inclusion feature in the syntax (while explaining the accusative syntactic diagnostics away in an independent manner). We do not follow this option here as cross-linguistic evidence shows that animacy or DOM-animacy systems do not obligatorily surface with LOC (oblique) morphology (across Algonquian, Bantu, etc., agreement can be used and not locative morphology). Interactions with perspectivization/sentience, on the other hand, are common, as animates are the preferred perspectival-centers.[15]

---

14. Although we cannot explain here why animacy requires licensing in a low position (below vP), while the accusative clitic (double) must be above v, similar observations have been made with respect to animate objects beyond Romance (López 2012).

15. As is well known, there are contexts where DOM overrides its canonical animacy/specificity features. For example, many types of nominal ellipsis present evidence of complex DP/KP structures (for example, in Romanian, see especially Cornilescu 2000 or Irimia 2020, a.o.), where

Isolating sentience as a separate feature on nominals (see also Kuno & Kaburaki's (1977) notion of empathy) can also explain why oblique DOM can be insensitive to (more canonical) information-structure specifications. An important line of research connects DOM to types of (secondary, low) topics (Iemmolo 2011, a.o.). The problem is that across Romance, oblique DOM does not necessarily individuate topics (from foci). For example, in Romanian, Spanish (López 2012, a.o.), Catalan, etc., oblique DOM is insensitive to information structure understood in these terms, while still showing sensitivity to animacy.

Thus, the main hypothesis is that the oblique DOM examined here[16] signals a type of accusative which needs the presence of an additional licenser below vP (the α projection).[17] This exact structural domain is also the locus of DAT or structural locative projections. Therefore, it is not surprising to obtain both DOM-LOC (Romanian), as well as DOM-LOC-DAT (Western Romance) syncretism. Crucially, an Enriched Case Hierarchy with two accusatives[18] as in (15) and (18) can capture all these patterns without incurring *ABA violations (that is, without syncretism insertion rules that skip cells in a hierarchy, see Caha 2009 for extensive discussion).

(18)

NOM=ACC syncretism   ACC= DAT syncretism with clitics

NOM > ACC$_S$ > DAT > GEN > LOC > ACC$_B$ > DAT > ABL/INS > …

Romanian LOC = DOM syncretism   Western Romance LOC = DOM = DAT syncretism
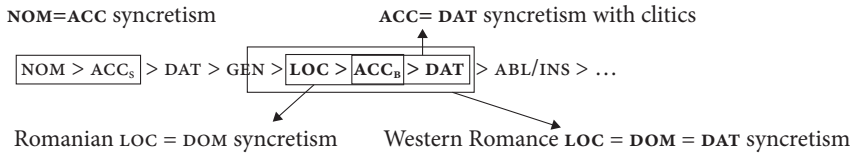
4. Conclusions

The account proposed here reconciles both the observation that DOM has the syntax of a structural accusative and its oblique appearance. The latter results from an additional licenser which is recruited for features (such as sentience) left behind by the initial licenser in the same domain. As it is a structural accusative at its core,

---

there can be various other PERSON/discourse-linking features which need licensing in the syntax, independently from [uC]. The unification point is, thus, that the α projection acts as licenser for *any* relevant features beyond [uC], which require licensing inside vP.

16. Note that we are not claiming that *all* types of (oblique) DOM (across Romance) involve additional licensing.

17. Here we leave aside the question whether DOM also involves (overt or covert) scrambling to this position.

18. For Starke (2017), the accusative closer to NOM is a 'smaller' (s) accusative, while the second accusative is 'bigger' (B) in terms of structure.

DOM does not have an oblique syntax, but its larger structure has syntactic and morphological consequences, that might trigger homomorphism with locatives at PF. In this respect, the present account also diverges from purely morphological resolutions of the DOM-OBL syncretism; in the latter, other syncretism patterns accusatives are part of are not easily derived.

Acknowledgements

Abbreviations

ABL	ablative
ACC	accusative
AUX	auxiliary
CL	clitic
DAT	dative
DEF	definite
DOM	differential object marking
GEN	genitive
F	feminine
IMP	impersonal
INS	instrumental
LOC	locative
M	masculine
MP	medio-passive
NOM	nominative
OBL	oblique
PASS	passive
PL	plural
PPA	past participle agreement
SE	Romance medio-passive/impersonal/reflexive marker
SG	singular
1/2/3	person

References

Anagnostopoulou, Elena. 2007. Clitic doubling. In Martin Everaert & Henk van Riemsdjik (eds.), *The Blackwell companion to syntax*, 519–581. Malden, MA: Blackwell Publishing.

Bárány, András. 2018. DOM and datives. *Glossa*. 3(1). 97.1–40. https://doi.org/10.5334/gjgl.639

Belletti, Adriana. 2005. Extended doubling and the VP periphery. *Probus* 17 (1): 1–35. https://doi.org/10.1515/prbs.2005.17.1.1

Belletti, Adriana. 2017. (Past) participle agreement. In Martin Everaert & Henk van Riemsdijk (eds.), *Wiley Blackwell companion to syntax*. Malden, MA: Blackwell. https://doi.org/10.1002/9781118358733.wbsyncom081

Bossong, Georg. 1991. Differential object marking in Romance and beyond. In Dieter Wanner & Douglas A. Kibbee (eds.), *New analyses in Romance linguistics*, 143–170. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.69.14bos

Caha, Pavel. 2009. The nanosyntax of case. Tromsø, Norway: University of Tromsø dissertation. http://ling.auf.net/lingbuzz/000956

Cornilescu, Alexandra. 2000. Notes on the interpretation of the prepositional accusative in Romanian. *Bucharest Working Papers in Linguistics* 1(2). 91–106.

D'Alessandro, Roberta & Ian Roberts. 2010. Past participle agreement in Abruzzese: Split-auxiliary selection and the null subject parameter. *Natural Language and Linguistic Theory* 28(1). 41–72. https://doi.org/10.1007/s11049-009-9085-1

Dobrovie-Sorin, Carmen. 1994. *The syntax of Romanian. Comparative studies in Romance*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110886597

Dobrovie-Sorin, Carmen. 1998. Impersonal *se* constructions in Romance and the passivization of unergatives. *Linguistic Inquiry* 29 (3). 399–437. https://doi.org/10.1162/002438998553806

Harðarson, Gísli Rúnar. 2016. A case for a Weak Case contiguity hypothesis: A reply to Caha. *Natural Language and Linguistic Theory* 34. 1329–1343. https://doi.org/10.1007/s11049-016-9328-x

Iemmolo, Giorgio. 2011. Towards a typological study of differential object marking and differential object indexation. Pavia, Italy: University of Pavia dissertation.

Irimia, Monica Alexandrina. 2020. Types of structural objects. Some remarks on differential object marking in Romanian. In András Bárány & Laura Kalin (eds.), *Case, agreement, and their interactions: New perspectives on Differential Object Marking*. Berlin: Mouton de Gruyter.

Irimia, Monica Alexandrina & Anna Pineda. 2019. Differential objects and datives: A homogeneous class? In Monica Alexandrina Irimia & Anna Pineda (eds.), *Lingvisticae Investigationes – Special issues on Differential objects and datives – a homogeneous class?* 42 (1), 1–6. https://doi.org/10.1075/li.00026.iri

Jaeggli, Osvaldo. 1982. *Topics in Romance syntax*. Dordrecht: Foris.

Jones, Michael Allan. 1993. *Sardinian syntax*. London: Routledge.

Kalin, Laura. 2018. Licensing and differential object marking: The view from Neo-Aramaic. *Syntax* 21(2). 112–159. https://doi.org/10.1111/synt.12153

Keine, Stefan. 2010. *Case and agreement from fringe to core: A minimalist approach*. Berlin: de Gruyter. https://doi.org/10.1515/9783110234404

Keine, Stefan & Gereon Müller. 2008. Differential argument encoding by impoverishment. In Marc Richards & Andrej L. Malchukov (eds.), *Scales* 83–136. Linguistische Arbeits Berichte 86. Universität Leipzig.

Kuno, Susumo & Etsuko Kaburaki. 1977. Empathy and syntax. *Linguistic Inquiry* 8(4). 627–672. https://www.jstor.org/stable/4178011?seq=1#page_scan_tab_contents

Ledgway, Adam. 2000. *A comparative syntax of the dialects of southern Italy: A minimalist approach*. Oxford: Blackwell.

López, Luis. 2012. *Indefinite objects: Scrambling, choice functions and differential marking*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9165.001.0001

Loporcaro, Michele. 2010. The logic of past participle agreement. In Roberta D'Alessandro, Adam Ledgeway & Ian Roberts (eds.), *Syntactic variation: the dialects of Italy*, 225–243. Cambridge: Cambridge University Press.

Manzini, M. Rita & Ludovico Franco. 2016. Goal and DOM datives. *Natural Language and Linguistic Theory* 34(1). 197–240. https://doi.org/10.1007/s11049-015-9303-y

Manzini, M. Rita & Ludovico Franco. 2019. 'Agreement of structural obliques' parameter. In Monica Alexandrina Irimia & Anna Pineda (eds.), *Lingvisticae Investigationes, Special issues on Differential objects and datives – a homogeneous class* 42(1), 82–101. https://doi.org/10.1075/li.00030.man

Manzini, M. Rita & Leonardo M. Savoia. 2005. *I dialetti italiani e romanci: Morfosintassi generativa*. 3 volumes. Alessandria: Edizioni dell'Orso.

Mardale, Alexandru. 2008. Microvariation within Differential Object Marking. *Revue Roumaine de Linguistique* 53(4). 448–467.

Martín, Juan. 2005. Aspectual quantization and [±] accusative case checking in Romance. In Twan Geerts, Ivo van Ginneken & Haike Jacobs (eds.), *Romance languages and linguistic theory 2003*, 177–196. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.270.11mar

Ormazabal, Javier and Juan Romero. 2013. Differential object marking: case and agreement. *Borealis: an International Journal of Hispanic Linguistics* 2.2:221–239. https://doi.org/10.7557/1.2.2.2808

Pancheva, Roumyana & María Luisa Zubizarreta. 2018. The Person Case Constraint. The syntactic encoding of perspective. *Natural Language and Linguistic Theory* 36(4). 1291–1337. https://link.springer.com/article/10.1007%2Fs11049-017-9395-7

Rodrîguez-Mondoñedo, Miguel. 2007. The syntax of objects. Agree and differential object marking. Mansfield, CT: University of Connecticut dissertation.

Săvescu Ciucivara, Oana. 2009. A syntactic analysis of pronominal clitic clusters in Romance: The view from Romanian. New York City, NY: New York University dissertation.

Starke, Michal. 2017. Resolving DAT=ACC (≠ GEN). *Glossa* 2(1). 104. 1–8. https://doi.org/10.5334/gjgl.408

Tigău, Alina M. 2011. *Syntax and interpretation of the direct object in Romance and Germanic languages with an emphasis on Romanian, German, Dutch and English*. Bucharest: Editura Universității București.

CHAPTER 5

A deletion account of referential null objects in Basque Spanish

Almike Vázquez-Lozares
University of Illinois at Urbana-Champaign

I propose a deletion analysis for null D(irect) O(bject)s in the Spanish of the Basque Country, based on a D(efiniteness)-feature on DO clitics and on Basque-Spanish *v*. When the DO clitic has a D-feature, as is the case with *lo(s)/la(s)*, its features form a proper subset of the features of *v*, and deletion of the clitic ensues in terms of a chain reduction à la Nunes (2004), as proposed in Roberts (2010). Importantly, DOs with human antecedents are not deleted because they undergo *leísmo*. I argue that *le* does not have a D-feature and, therefore, its features are not a proper subset of the features of *v*, thus precluding deletion.

Keywords: null objects, *leísmo*, clitics, agreement, definiteness-feature, chain reduction, Spanish, Basque, contact

1. Introduction

Null objects have been one of the most widely discussed features in the variety of Spanish spoken in the Basque Country (hereafter Basque Spanish or B-Spanish) (Landa 1995; Urrutia Cárdenas 2003; Franco & Landa 2003; Gómez Seibane 2011, 2012; Camus Bergareche & Gómez Seibane 2015; Sainz-Maza Lecanda & Schwenter 2017, *inter alia*). Referential null objects, illustrated in (1a), refer to the missing direct object (DO) clitic that typically replaces DO DPs in most Spanish varieties, as in (1b).

(1) a. B-Spanish
         Ya   ø   he   visitado.
         Already DO   have visited
         "I have already visited it."
     b. Standard Spanish
         Ya   lo   he   visitado.
         Already DO   have visited

Previous studies of Basque Spanish null objects agree that definiteness, specificity and, more relevantly, inanimacy of the antecedent are the key factors that allow DO clitics *lo(s), la(s)* to be dropped. Based on data from Person Case Constraint contexts, I argue that a D(efiniteness)-feature is the relevant factor that determines the possibility of null objects: because of a D-feature found in B-Spanish *v* and third-person accusative clitics, when the object agrees with *v* and all of their features match, the lowest copy of the clitic can be deleted in terms of chain reduction à la Nunes (2004) and as initially proposed by Roberts (2010).

The paper is organized as follows: Section 2 presents the distribution of null objects in B-Spanish and shows that null DOs must be third person objects that agree with *v*. Section 3 puts forward an analysis based on Roberts's (2010) proposal that languages that allow null referential objects have a D-feature in *v*. I argue that this D-feature is present in Basque *v*, and gets transferred to B-Spanish. Since third person accusative clitics also have a D-feature, Agree between *v* and the DO allows third person null DOs. Section 4 concludes the paper with discussion of other varieties of Spanish that allow null objects, as well as discussion on how Basque may affect Spanish.

2. The distribution of referential null objects in B-Spanish

The phenomenon of referential null objects in Spanish refers to the omitted third person DO (accusative) clitics *lo(s), la(s)*. Arbitrary null DOs are available with indefinite, unspecific antecedents such as in (2), in all varieties of Spanish, including Basque- and non-Basque Peninsular Spanish (NB-Spanish). However, these semantic restrictions do not seem to apply in B-Spanish, where definite, specific, referential null objects are possible too, as in (3).

(2) A: *Quiero comprar unos pastelitos$_i$ para la fiesta.*
         Want buy   some cakes   for the party
     B: *Ya   ø$_i$ he   comprado yo.*
         Already ø have bought   I
         "–I want to buy some cakes for the party. –I already bought (some) myself."

(3) A: *¿Has comprado el   pollo$_j$?*
         Have bought   the chicken
     B: *Sí, ø$_j$ he   comprado.*   (* in NB-Spanish)
         Yes ø have bought
         "–Have you bought the chicken? –Yes, I have bought (it)."

Previous studies have extensively shown that specificity, definiteness and inanimacy (or non-humanness) of the antecedent are the key semantic factors that contribute to null objects in B-Spanish (Landa 1995; Landa & Franco 2000; Camus Bergareche & Gómez Seibane 2015, a.o.).

2.1   The relevance of case

A crucial assumption in previous literature has been that null objects can only take inanimate antecedents. Example (4) below, with an animate antecedent, is ungrammatical in B-Spanish.

(4)   A: *¿Dónde está Juan?*
              Where   is   Juan
      *B: *Ahora no sé,   pero ø he   visto antes.*
              Now   not know but ø have seen earlier
              "–Where is Juan? –I don't know where he is now, but I saw him earlier."

In B-Spanish, human DOs are pronominalized with the dative *le(s)* as opposed to the accusative *lo(s)/la(s)*, a phenomenon known as *leísmo*. As we will see in § 3, *leísmo* precludes the possibility of these objects being null.

One particular case in which *leísmo* is blocked in Peninsular varieties of Spanish pertains to the Person Case Constraint (PCC). In broad terms, the PCC bans certain sequences of clitics or agreement markers (first noted by Perlmutter 1971 and studied by Bonet 1991 for Spanish). In Spanish, the PCC blocks combinations of two clitics unless the second one is an accusative third person (*lo(s), la(s)*).[1] That is why when a dative indirect object (IO) clitic is already present, *leísmo* cannot take place, as shown in (5a) and (6a). In these cases, DOs need to be cliticized with the etymological accusative form *lo(s)/la(s)*, regardless of animacy.

(5)   a.   *A   Juan_j, me **le**_j   presentaron ayer.*
              DOM Juan IO DO presented   yesterday
      b.   *A   Juan_j, me **lo**_j   presentaron ayer.*
              DOM Juan IO DO presented   yesterday
              "Juan, they introduced him to me yesterday."

---

1.   Note that the PCC applies to clitics that refer to 'thematic arguments', but not to a combination containing an ethical dative (see Laka 1993: 28). Thus, (6a) is grammatical with a reading in which *me* is not an IO but is instead understood as affected by the action, as illustrated in (i) below.

(i)   *A   los niños, me les ha mandado María a Madrid.*
        DOM the kids   IO DO has sent   María to Madrid
        "The kids, María sent them to Madrid (on me)."

(6) a. *A   los niños*_i*, me **les**_i* ha mandado María.
         DOM the kids   IO DO has sent   María
     b.   A   los niños*_i*, me **los**_i* ha mandado María.
         DOM the kids   IO DO has sent   María
         "The kids, María sent them to me."

Interestingly, in these restricted cases in which human DOs have to be pronominalized with the accusative clitic, B-Spanish allows null objects too, as in (7) below.

(7) a.   A   Juan*_j*, me ø*_j* presentaron ayer.
         DOM Juan IO ø presented   yesterday
         "Juan, they introduced him to me yesterday."
     b.   A   los niños*_i* me ø*_i* ha mandado María.
         DOM the kids   IO ø has sent   María
         "The kids, María sent them to me."

Landa (1995: 132) proposes that, in these contexts, the human DO is lowered to a [- involved] status, as if it were "thought of as an object or depersonified". Landa's proposal allows us to keep animacy as the crucial feature for null objects. Here, I propose that the licensing of null objects in B-Spanish is not directly related to animacy of the antecedent, but to the features of the clitic and its agreement with *v* instead: when the DO clitic keeps all its features and, crucially, its D(efiniteness)-feature, the object can be null. Therefore, even if the antecedent is human, if *leísmo* cannot take place, the DO has the right features to be null.

3.   Analysis

In this section, I propose an analysis for referential null objects in B-Spanish. Following Franco & Landa (1991) and Landa (1995), I assume that B-Spanish null objects are *pro*. Next, I propose that this *pro* is licensed in object position thanks to a D(efiniteness)-feature in *v* which is transferred from Basque. Importantly, *pro* is only available with determiner type clitics *lo(s), la(s)*, since these are the only clitics that contain a D-feature.

Franco & Landa (1991), and Landa (1995) argue that B-Spanish null objects are *pro*, based on their behavior in a series of contexts: complex NPs, doubly filled complementizers, sentential subjects, adjunct islands, and contexts relevant to the Weak Crossover Constraint. For the sake of space, these arguments are not summarized here and the reader is directed to the original works. As for arbitrary null objects, I assume that they are variables, as argued in Campos (1986); Modesto (2000) and, for B-Spanish, in Landa & Franco (1992). Therefore, arbitrary null objects follow different constraints and licensing conditions which are beyond the scope of this chapter.

Arguing that referential null objects are *pro* does not require the positing of a new element in Spanish, since Spanish already contains *pro* for referential null subjects. B-Spanish is special in that it allows *pro* in object position, too. This difference in the distribution of *pro* in B- and NB-Spanish is arguably related to transfer from Basque. Basque has referential null subjects and objects, as illustrated in (8), where the auxiliary verb encodes the morphology of the arguments.

(8) *Bidali d- i-   zu-   t.*
    sent   L- PRES3SG- DAT2SG- ERG1SG[2]
    "I have sent it to you."

The availability of null arguments in Basque seems to affect B-Spanish. B-Spanish speakers allow referential null objects, which are only marginally accepted in NB-Spanish (see § 4 for further discussion of contact effects as well as of null objects in other varieties).

Interestingly, Spanish in contact with Quechua shows parallelisms with Spanish in contact with Basque in terms of allowing referential null objects. Both Quechua and Basque allow referential null objects, as we saw in (8) above for Basque and as shown in (9) below for Central Quechua, with data from Sánchez (1999: 234).

(9) *Manam rikura- ni-   chu*
    not   see.1s- 1SG- NEG
    "I did not see her/any."

Spanish in contact with Quechua also allows null objects, as illustrated in (10), from Escobar (1990: 89).

(10) *A veces en la noche dejo su   quacker ya   preparado, en la*
     At times in the night leave their oatmeal already prepared in the
     mañana ø calientan y   ø toman.
     morning ø heat.up and ø take
     "Sometimes I leave their oatmeal already prepared at night and in the morning they heat it up and they eat it."

This phenomenon is not attested in the Spanish in contact with other languages that do not allow referential null objects, such as Catalan, Galician, French, or English. This may suggest that referential null objects in B-Spanish (as well as in Spanish in contact with Quechua) are related to language contact.

---

2. L = 'left' or 'linearization-related', from Arregi & Nevins (2012), who propose that the morpheme *d-* is an epenthetic morpheme inserted postsyntactically due to a constraint on the Linearization of T.

3.1   The nature of accusative vs. dative clitics

As we saw in § 2, referential null objects are most common with non-human an-tecedents, i.e., with those that do not undergo *leísmo*, and that would require the accusative clitic *lo(s)/la(s)*[3] Interestingly, there is much literature on the distinction between third person accusative clitics *lo(s)/la(s)* as determiners, and the rest of accusative and dative clitics as agreement markers (Roca 1996; Uriagereka 1995; Torrego 1998; Bleam 2000; Franco 2000; Ormazabal & Romero 2013, among others).

The idea that clitics mark agreement goes back to Silva-Corvalán (1981); Borer (1984); Suñer (1988); Saltarelli (1989), and Sportiche (1996) amongothers. In particular, Franco & Mejías-Bikandi (1997); Franco (2000) and, more recently, Ormazabal & Romero (2013) show that B-Spanish has developed an agreement marker type of DO (*le/s*) that shows properties of the restof the agreement markers.

Ormazabal & Romero (2013) offer data involving the doubling of the quantifier *todos*, and data with negative quantifiers in clitic left dislocation constructions. In both of these contexts, *lo/la* pattern differently from the rest of the clitics. The assumption is that *lo/la* do not mark agreement, like the other clitics do (first and second persons, and third person dative clitics). Importantly, B-Spanish DO *le* patterns like the agreement markers and unlike *lo/la* (see Ormazabal & Romero (2013) for the full-fledged explanation of the data).

In other work, Ormazabal & Romero (2007) extend the PCC to the Object Agreement Constraint, which basically specifies that the verbal complex can only encode agreement for one argumental object. Observe that first and second person object markers cannot co-occur with an argumental dative, as in (11b): here, a sec-ond person IO marker co-occurs with a third person DO in 'dative' form, because of *leísmo*, and the combination is ungrammatical.

---

3.   The system described here only considers *animate leísmo*, which is the most common type of *leísmo* in B-Spanish, as well as the most studied one in the literature on the B-Spanish clitic system (see Franco 1993; Landa 1995; Urrutia Cárdenas 1995; Rodríguez-Ordóñez 2016, 2017; Fernandez-Ordóñez 1994, 1999; Ormazabal & Romero 2007, 2013; Odria 2019 to mention some). Note, however, that there are other types of *leísmo*, even in the Basque Country, which affect inanimate objects, as in (i) below. Rodríguez-Ordóñez (2017: 328, fn. 10) describes this use as rare and restricted to elderly people.

(i)   *El   teléfono,   le   he   dejado   en   la   mesa.*
       The phone   DO   have left   on   the table.
       "The phone, I have left it on the table."

This *leísmo* can be found in other areas of Spain as well. On the other hand, there is a different type of *leísmo* that affects animate masculine DOs, which is found in the central and northern area of Spain.

(11) a. *Le llevé a tu hijo a casa.*
          DO took DOM your son to home
          "I took your son home."

     b. *Te (* le) llevé (a) tu hijo a casa.*
          IO   DO took DOM your son to home
          "I brought you your son home."   (Ormazabal & Romero 2013: 317)

The repair in these contexts is to avoid *leísmo* by using the etymological accusative form *lo/la* instead.

(12) *Te lo llevé a casa.*
       IO DO took to home
       "I brought you it/him (home)."   (Ormazabal & Romero 2013: 317)

These data can easily be explained if *lo/la* are external to the agreement system, and they are actually of a different nature. The idea that I adopt here is that *lo/la* are determiner-type clitics, while all the other clitics mark agreement, including B-Spanish DO *le* (Ormazabal & Romero 2013). Furthermore, since the distribution of *lo/la* is constrained by the semantics of the DP they refer to (see Franco 1993; Roca 1996; Franco & Mejías-Bikandi 1997; and Ormazabal & Romero 2013, a.o.), I propose that *lo/la* have a D-feature which the other object markers lack.

3.2   A deletion analysis of null objects

In Romance Null Subject Languages (NSLs), a D-feature related to rich agreement is said to be responsible for null subjects (Chomsky 1995; Alexiadou & Anagnostopoulou 1998; Holmberg 2005; Roberts 2010, *inter alia*). In general terms, a D-feature in T satisfies the EPP feature, thus making overt subjects unnecessary in NSLs.

Specifically, Roberts (2010) argues that deletion of subjects in NSLs is the result of a chain reduction: T has a D-feature, as do pronouns; when T and the pronoun Agree for phi-features and case, the features of the pronoun form a proper subset of the features of T, and only the highest copy is pronounced (T), making the subject null.

Roberts extends this analysis to languages which allow null objects, such as Pashto, by proposing that *v* has a D-feature that works the same way. Basque is another language in which this analysis would work. As we saw in Example (8), repeated here as (13), Basque is a pro drop language that shows agreement with its arguments in the auxiliary verb.

(13) *Bidali d- i- zu- t.*
       sent   L- PRES3sg- DAT2sg- ERG1sg
       "I have sent it to you."

In Basque, the availability of null arguments is related to rich agreement. Following Roberts, rich agreement that allows null objects involves a D-feature in *v*. Arguably, as a result of transfer, this D-feature in *v* is available in the grammar of B-Spanish speakers as well and allows null *pro* in object position. While *pro* is already part of the grammar of Spanish, what is transferred from Basque is the property to have *pro* in object position: this property is formalized in terms of a D-feature in *v*. Consequently, I argue that *v* is the locus of variation, ashas already been proposed in previous work such as Torrego (1998); González-Vilbazo & López (2012); Ordoñez & Roca (2019), etc.

Recall that DO clitics *lo(s), la(s)* contain a D-feature while B-Spanish *le(s)* and the rest of IO clitics do not. When *v* Agrees with the DO clitic, *v* gets its phi-features valued and the clitic gets case. Since both *v* and the clitic *lo(s), la(s)* have a D-feature, Agree results in the features of the clitic forming a proper subset of the features of *v*.

(14)
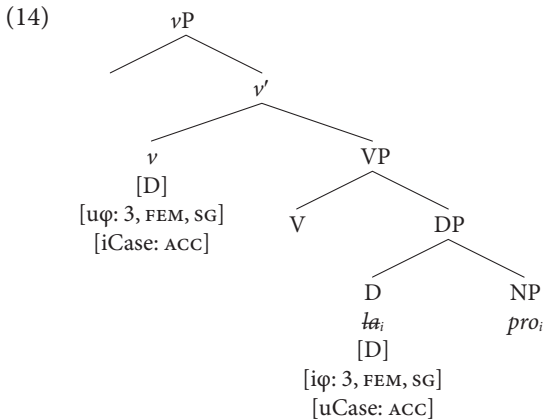
Following Roberts (2010), deletion of the clitic takes place via chain reduction in the sense of Nunes (2004). Since the goal of Agree constitutes an exact copy of the features of the probe, all identical copies delete except for the highest one.

Agreement markers, on the other hand, do not have a D-feature, which prevents them from undergoing deletion. Since B-Spanish DO *le* is an agreement marker, deletion of animate DO clitics does not take place because the features of *le* do not form a proper subset of the features of *v*.

(15) a. *Le   vi   (en la   fiesta).*
          DO.DAT saw   (at the party).
          "I saw him/her at the party."

b.

vP
└ v'
  ├ v
  │ [D]
  │ [uφ: 3, FEM, SG]
  │ [iCase: ACC]
  └ VP
    ├ V
    └ DP
      ├ D
      │ le_i
      │ [iφ: 3, FEM, SG]
      │ [uCase: ACC]
      └ NP
         pro_i
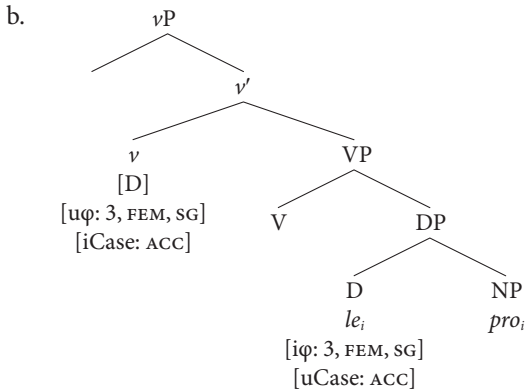
It is *leísmo*, the lack of a D-feature in the clitic, which causes the inability of animate DOs to be null.

As we saw in § 2, in PCC contexts, animate DOs can be null. This is explained in this analysis because *leísmo* is blocked in PCC contexts, that is, the clitic does not lose its D-feature, as shown in (16). Consequently, the clitic with all its features agrees with v and the exact matching of their features allows its deletion, regardless of the animacy of the antecedent.

(16)   a.   *A   Juan$_i$, te   ø$_i$ presentaron en la   fiesta.*
            DOM Juan   IO ø   introduced at   the party
            "Juan, they introduced him to you at the party."

b.

vP
└ v'
  ├ v
  │ *presentaron*
  │ D
  │ [uφ: 3, MASC, SG]
  │ [iCase: ACC]
  └ ApplP
    ├ pro_i
    └ Appl'
      ├ ApplP
      │ te_i
      └ VP
        ├ v
        │ *presentaron*
        └ DP
          ├ D
          │ ~~lo~~$_h$
          │ [D]
          │ [iφ: 3, MASC, SG]
          │ [uCase: ACC]
          └ NP
             pro_h
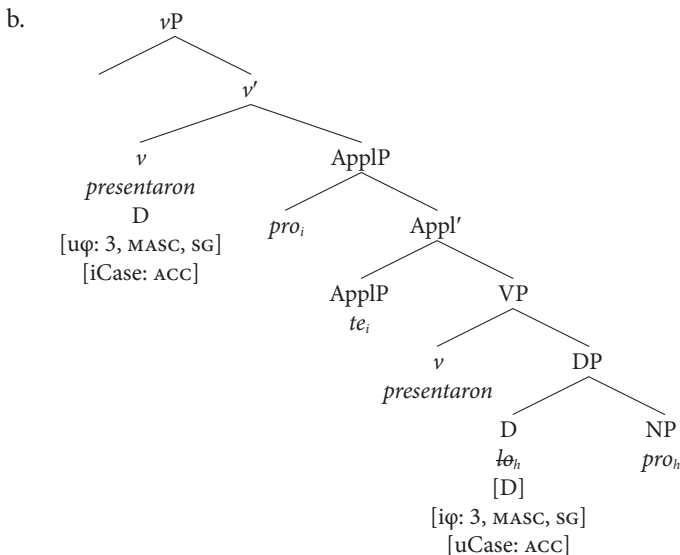
4.   Final remarks

In this paper I have proposed that a D-feature in *v* is available in B-Spanish as a result of contact with Basque. However, referential null objects are, in fact, also attested in monolingual varieties. Null objects are licensed in certain constructions in all varieties of Spanish, when the referent is recoverable from the context (Masullo 2003), as shown in (17).

(17)   Context: Two persons leaving a room, one says to the other:
    Apaga   ø   [i.e., la   luz,   la   televisión, etc.]
    Turn.off DO [i.e., the light, the television, etc.]
    "Turn it off."   (Alamillo & Schwenter 2007: 113)

Monolingual varieties may also allow null DOs with propositional antecedents: Alamillo & Schwenter (2007) find that, in Madrid, these null DOs are restricted to expressions like *no sé* "I don't know" and to non-declarative sentences; in Mexico City, null DOs are not restricted to those contexts but they are dependent on other factors such as the presence of *ya* "already", the presence of modal adverbials, and the person (1st and 2nd vs. 3rd), amongst others. More relevantly, there are other monolingual varieties, such as Rioplatense Spanish, which have referential null DOs across the board (Masullo 2003; Schwenter 2006; Maddox 2019).

    Maddox (2019) proposes that referential null DOs are licensed in Rioplatense by a D-feature in *v*, like I proposed here for B-Spanish. Maddox builds upon van Gelderen's (2011) Object Agreement Cycle, and argues that the D-feature is present in *v* as a result of the reanalysis of clitics that is part of the cycle.

    While I propose here that the availability of null objects in B-Spanish is influenced by contact from Basque, it could be the case that contact is simply accelerating a natural process (or cycle) in Spanish. Thus, non-Basque Peninsular Spanish has not reached the stage at which null objects are allowed, but B-Spanish has, due to contact. On the other hand, the idea put forth here that the D-feature in *v* is directly borrowed from Basque is justified by the fact that bilingual speakers always try to maximize the "common ground"[4] (Filipović & Hawkins 2019: 1229) of both their languages. By borrowing the D-feature from Basque *v* the structure in B-Spanish aligns with the structure in Basque, where null objects are common.

---

4.   Filipović & Hawkins (2019) propose five general principles that underlie bilingual speakers' language behavior, including maximizing common ground. This means that if the two languages share a given construction, this shared construction will be used more frequently in both languages, even if that means using a structure that is not the preferred or majority one in one of the languages (i.e., one of the two languages might have a more common structure to express the same). In this case, monolingual Spanish has overt clitics *lo, la*, which are more common than null objects. However, a bilingual Basque-Spanish speaker prefers the null object option because it is common to both languages.

In fact, Arregi & Nevins (2012) argue that Basque auxiliary verbs are formed by clitics which the arguments project. All arguments generate clitics except for third person absolutive arguments, which is the case of direct objects. Thus, there is no third person DO clitic in Basque, which coincides with the availability of null objects in B-Spanish.

To conclude, note that B-Spanish speakers still allow overt objects *lo(s),* *la(s),* since these are grammatical in the NB-Spanish grammar. This suggests that B-Spanish speakers have access to the structure with overt object clitics found in NB-Spanish, where the assumption is that *v* has no D-feature, as well as to the structure with null objects, licensed by the D-feature transferred from Basque.

References

Alamillo, Assela Reig & Scott. A. Schwenter. 2007. Null objects and neuter *lo*: A cross-dialectal variationist analysis. In Jonathan Holmquist, Gerardo Augusto Lorenzino & Lotfi Sayahi (eds.), *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics*, 113–121. Somerville, MA: Cascadilla Proceedings Project.

Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parametrizing Agr: Word order, V-movement and EPP-checking. *Natural Language and Linguistic Theory* 16(3). 491–539. https://doi.org/10.1023/A:1006090432389

Arregi, Karlos & Andrew Nevins. 2012. *Morphotactics: Basque auxiliaries and the structure of spellout* (Studies in Natural Language and Linguistic Theory 86). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-3889-8

Bleam, Tonia M. 2000. Leísta Spanish and the syntax of clitic doubling. IRCS Technical Reports Series.

Bonet, Eulalia. 1991. Morphology after syntax: Pronominal clitics in Romance languages. Cambridge, MA: MIT dissertation.

Borer, Hagit. 1984. *Parametric syntax*. Dordrecht: Foris. https://doi.org/10.1515/9783110808506

Campos, Héctor. 1986. Indefinite object drop. *Linguistic Inquiry* 17. 354–359.

Camus Bergareche, Bruno & Sara Gómez Seibane. 2015. Nuevos datos sobre la omisión de objetos en el castellano del País Vasco. *Círculo de Lingüística Aplicada a la Comunicación* 6. 211–236.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Escobar, Anna María. 1990. *Los bilingües y el castellano en el Perú*. Lima, Perú: Instituto de estudios peruanos.

Fernández-Ordóñez, Inés. 1994. Isoglosas internas del castellano. El sistema referencial del pronombre átono de tercera persona. *Revista de Filología Española* 74(1/2). 71–125. https://doi.org/10.3989/rfe.1994.v74.i1/2.499

Fernández-Ordóñez, Inés. 1999. Leísmo, laísmo y loísmo. In Ignacio Bosque and Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 1317–1398. Madrid: Espasa Calpe.

Filipović, Luna & John A. Hawkins. 2019. The Complex Adaptive System Principles model for bilingualism: Language interactions within and across bilingual minds. *International Journal of Bilingualism* 23(6). 1223–1248. https://doi.org/10.1177/1367006918781076

Franco, Jon. 1993. Conditions on clitic doubling: The Agreement Hypothesis. *ASJU* 27. 285–298.

Franco, Jon. 2000. On object agreement in Spanish. Los Angeles, CA: University of Southern California dissertation.

Franco, Jon & Alazne Landa. 1991. Spanish object clitics as verbal agreement morphemes. *MIT Working Papers in Linguistics*, 14. 99–114. Cambridge, MA.

Franco, Jon & Alazne Landa. 2003. Null objects revisited. In Rafael Núñez-Cedeño, Luis López & Richard Cameron (eds.), *A Romance perspective on language knowledge and use: Selected papers from the 31st Linguistic Symposium on Romance Languages*, 311–326. Amsterdam, Philadelphia: Benjamins. https://doi.org/10.1075/cilt.238.23fra

Franco, Jon & Errapel Mejías-Bikandi. 1997. Overt and covert raising to [Spec, AGR$_{DO}$] and the interpretation of objects. *Linguistic Analysis* 27(1–2). 79–107.

Van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199756056.001.0001

Gómez Seibane, Sara. 2011. Objetos nulos en hablantes bilingües de vasco y español. *Oihenart* 26. 259–276.

Gómez Seibane, Sara. 2012. La omisión y duplicación de objetos en el castellano del País Vasco. In Bruno Camus Bergareche & Sara Gómez Seibane (eds.), *El castellano del País Vasco*, 193–214. Bilbao: UPV.

González-Vilbazo, Kay & Luis López. 2012. Little *v* and parametric variation. *Natural Language and Linguistic Theory* 30. 33–77. https://doi.org/10.1007/s11049-011-9141-z

Holmberg, Anders. 2005. Is there a little pro? Evidence from Finnish. *Linguistic Inquiry* 36. 533–564. https://doi.org/10.1162/002438905774464322

Laka, Itziar. 1993. The structure of inflection: A case study in X0 syntax. In José Ignacio Hualde & Jon Ortiz de Urbina (eds.) *Generative studies in Basque linguistics*, 21–70. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.105.02lak

Landa, Alazne. 1995. Conditions on null objects in Basque Spanish and their relation to leísmo and clitic doubling. Los Angeles, CA: University of Southern California dissertation.

Landa, Alazne & Jon Franco. 1992. Objetos nulos en el castellano del País Vasco: dos estatus para dos interpretaciones. *ASJU* 26(3). 777–792.

Landa, Alazne & Jon Franco. 2000. The role of syntax in language variation. Paper presented at the International Conference on Language Variation in Europe (ICLaVE), Barcelona, Spain, June 29-July1.

Maddox, Matthew L. 2019. Cycles of agreement: Romance clitics in diachrony. Urbana, IL: University of Illinois at Urbana-Champaign dissertation.

Masullo, Pascual José. 2003. Clitic-less definite object drop in River Plate Spanish. Paper presented at the Linguisitc Symposium on Romance Languages (LSRL) 33, Bloomington, IN, April 24–27.

Modesto, Marcello. 2000. On the identification of null arguments. Los Angeles, CA: University of Southern California dissertation.

Nunes, Jairo. 2004. *Linearization of chains and sideward movement*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/4241.001.0001

Odria, Ane. 2019. Differential Object Marking in Basque and Spanish dialects. In Ane Berro, Beatriz Fernández & Jon Ortiz de Urbina (eds.), *Basque and Romance. Aligning grammars*, 243–275. Boston: Brill. https://doi.org/10.1163/9789004395398_009

Ordóñez, Francisco & Francesc Roca. 2019. Differential Object Marking (DOM) and clitic subspecification in Catalonian Spanish. In Ángel Gallego (ed.), *The syntactic variation of Spanish dialects*, 35–59. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190634797.003.0002

Ormazabal, Javier & Juan Romero. 2007. The Object Agreement Constraint. *Natural Language and Linguistic Theory* 25(2). 315–47. https://doi.org/10.1007/s11049-006-9010-9

Ormazabal, Javier & Juan Romero. 2013. Object clitics, agreement and dialectal variation. *Probus* 25(2). 301–344. https://doi.org/10.1515/probus-2013-0012

Perlmutter, David. 1971. *Deep and surface structure constraints in syntax*. New York: Rinehart and Winston Inc.

Roberts, Ian. 2010. A deletion analysis of null subjects. In Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.), *Parametric variation: Null subjects in Minimalist Theory*. Cambridge: Cambridge University Press.

Roca, Francesc. 1996. Morfemas objetivos y determinantes: Los clíticos del español. *Verba* 23. 83–119.

Rodríguez-Ordóñez, Itxaso. 2016. DOM in Basque: Grammaticalization attitudes and ideological representations. Urbana, IL: University of Illinois at Urbana-Champaign dissertation.

Rodríguez-Ordóñez, Itxaso. 2017. Reexamining Differential Object Marking as a linguistic contact-phenomenon in Gernika Basque. *Journal of Language Contact* 10. 318–352. https://doi.org/10.1163/19552629-01002004

Sainz-Maza Lecanda, Lorena & Scott A. Schwenter. 2017. Null objects with and without bilingualism in the Portuguese- and Spanish-speaking world. In Kate Bellamy, Michael W. Child, Paz González, Antje Muntendam & M. Carmen Parafita Couto (eds.), *Multidisciplinary Approaches to Bilingualism in the Hispanic and Lusophone World*, 95–119. Amsterdam: John Benjamins. https://doi.org/10.1075/ihll.13.05sai

Saltarelli, Mario. 1989. Syntactic shift and the creation of clitics in Romance. In Carl Kirschner & Janet DeCesaris (eds.), *Studies in Romance linguistics. Selected papers from the Seventeenth Linguistic Symposium on Romance Languages (LSRL 17), Rutgers University, 27–29 March 1987*, 347–364. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.60.21sal

Sánchez, Liliana. 1999. Null objects and Dº features in contact Spanish. In J. Marc Authier, Barbara E. Bullock & Lisa A. Reed (eds.), *Formal Perspectives on Romance Linguistics*, 227–243. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.185.17san

Schwenter, Scott A. 2006. Null objects across South America. In Timothy L. Face & Carol A. Klee (eds.), *Selected Proceedings of the 8th Hispanic Linguistics Symposium*, 23–36. Somerville, MA: Cascadilla Proceedings Project.

Silva-Corvalán. 1981. The diffusion of object-verb agreement in Spanish. In Heles Contreras & Jürgen Klausenburger (eds.), *Proceedings of the Tenth Anniversary Symposium on Romance Linguistics, Papers in Romance. Volume 2 of Papers in Romance: Supplement*, 163–176. Seattle, WA: University of Washington.

Sportiche, Dominique. 1996. Clitic constructions. In Johan Rooryck & Laurie Zaring (eds.), *Phrase structure and the lexicon. Studies in Natural Language and Linguistic Theory, Volume 33*, 213–276. Dordrecht: Springer. https://doi.org/10.1007/978-94-015-8617-7_9

Suñer, Margarita. 1988. The role of agreement in clitic doubled constructions. *Natural Language and Linguistic Theory* 6. 391–434. https://doi.org/10.1007/BF00133904

Torrego, Esther. 1998. *The dependencies of objects. Linguistic Inquiry Monographs 34*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/2337.001.0001

Uriagereka, Juan. 1995. Aspects of the syntax of clitic placement in Western Romance. *Linguistic Inquiry* 26. 79–124.

Urrutia Cárdenas, Hernán. 1995. Morphosyntactic features in the Spanish of the Basque Country. In Carmen Silva-Corvalán (ed.), *Spanish in four continents: Studies in language contact and bilingualism*, 243–259. Washington, DC: Georgetown University Press.

Urrutia Cárdenas, Hernán. 2003. Los clíticos de tercera persona en el País Vasco. *Cauce, Revista de Filología y su Didáctica* 26. 517–538.

Same EPP, different null subject type

Julianne Doner and Çağrı Bilgin
University of Toronto

It is commonly assumed that all languages with Extended Projection Principle (EPP) type X will have null subject language (NSL) type Y, and vice versa (see Holmberg 2005, *inter alia*). We argue that although EPP type and NSL type interact, they are *not* co-extensive. We demonstrate that General Spanish, Brazilian Portuguese, and French have *different* NSL types (consistent, partial, and non-NSLs, respectively), but share the *same* EPP type (DP EPP). We also argue that both Brazilian Portuguese and Dominican Spanish underwent a change in EPP type, followed by the loss of agreement and a change in NSL type. Crucially, EPP type and NSL type did *not* change simultaneously. It therefore follows that EPP and NSL type cannot be co-extensive.

Keywords: null-subject languages, EPP, loss of agreement, Spanish, Brazilian Portuguese, French

1. Introduction

Since Alexiadou & Anagnostopoulou (1998), it is often assumed that the Extended Projection Principle (EPP) type and the null subject language (NSL) type of a language are largely co-extensive (e.g., as discussed in Holmberg 2005); that is, all languages with EPP type X will have NSL type Y, and vice versa. However, there is no consensus on how null subjects (NSs) are licensed, nor is there consensus on what type of EPP some NSLs have. One reason for this may be that a correlation between NSL type and EPP type is often assumed. We show that although EPP type and NSL type interact, they are *not* co-extensive, since several languages with *different* NSL types have the *same* EPP type (§ 2), and since EPP type and NSL type can vary independently, both synchronically and diachronically (§ 3).

1.1   Background on the EPP

The EPP was proposed by Chomsky (1981, 1982) to explain why subjects are obligatory in English. In this chapter, we define the EPP as the obligatory movement of some element into the inflectional domain. Different cross-linguistic varieties of the EPP have been proposed (cf. Biberauer 2010), including a contrast between varieties that can be checked by either a phrase or a head (Alexiadou & Anagnostopoulou 1998), by either a nominal or a verbal element (Massam & Smallwood 1997; Davies & Dubinsky 2001), or by pied-piping the entire vP (Richards & Biberauer 2005). Based on a comparison of these types, Doner (2019) proposes the EPP typology in Table 1.

Table 1. A typology of the EPP (Doner 2019: 79)

		Pied-Piping	No Pied-Piping
Argument-EPP	**X°-EPP**	(a) German, Icelandic	(b) Greek, Italian
	XP-EPP	(c) Dutch, Afrikaans	(d) English, French, Finnish
Predicate-EPP	**X°-EPP**	(e) Celtic (except Breton)	(f) Inuktitut
	XP-EPP	n/a	(g) Niuean

Two EPP types are relevant for this chapter, D-on-V EPP languages and DP EPP languages. In D-on-V EPP languages, the EPP is checked by a [D] feature on the verb, realized as rich agreement, when the verb raises to T. The [D] feature indexes an argument and raises to T through head movement without pied-piping, so it is type (b) in Table 1. In these languages, there must be rich agreement and there must be V-to-T raising in clauses of all types (including non-finite clauses). On the other hand, in DP EPP languages, such as English, the EPP is checked by the raising of some DP. This DP is an argument which raises through phrasal movement without pied-piping, and so it is type (d) in Table 1. In DP EPP languages, there may or may not be V-to-T movement (if there is, it is not triggered by the EPP), expletives are possible, and subjects are usually overt, but may be null in some languages (Holmberg 2005). However, NSs in DP EPP languages must be syntactically active (e.g., able to bind anaphors) (Holmberg, Nayudu & Sheehan 2009).[1]

---

1.   The EPP properties of a language seem to change in some registers of casual speech and in some written registers. This seems to be through a process of deletion, motivated either syntactically (Haegeman 2000, 2013) or prosodically (Weir 2012). We set these cases aside.

1.2   Background on NSLs

Null Subject Languages (NSLs) can be divided into consistent and partial NSLs, among others (see, e.g., Barbosa 2011). Consistent NSLs, such as European Portuguese, allow NSs in almost any person, while partial NSLs, such as Finnish and Brazilian Portuguese (BP), do not allow referential third person NSs in matrix clauses, but do allow NSs in first and typically second persons.[2] In contrast, non-NSLs, such as French[3] and English, do not allow NSs in any person (save a few exceptions).

It is important to note that as languages can vary greatly with regard to how many NSs they allow (i.e., which persons), the NSL status of a language is on a spectrum with non-NSLs that do not allow any NSs on one end of the spectrum and completely consistent NSLs that always allow NSs. For instance, BP only allows first person NSs, while Finnish allows first and second person NSs, although they are both classified as partial NSLs. In discussing 'NSL types' below, we refer to the typology/spectrum outlined here.

2.   Different NSL type, same EPP

As shown in Table 2, different languages share the same EPP type but differ in NSL type. We begin by demonstrating the different NSL types of French, BP and Spanish, and then arguing that they all share the same DP-EPP type.

Table 2. EPP type compared to NSL type

	NSL type	EPP type	V to T?	Selected citations
French	non	DP EPP	Yes	Roberts 2010
Brazilian Portuguese (BP) Finnish	partial	DP EPP	Yes	Barbosa 2009; Holmberg 2005
General Spanish (GS)	consistent	DP EPP	Yes	Toribio 2000; Goodall 2001

---

2.   Partial NSLs allow third person NSs in embedded clauses with matrix antecedents.

3.   Roberts (2010: 311) classifies French as a "(very) partial NSL"; however, the contexts in which NSs are licensed in French are so highly restricted that, for our purposes, it is in essence a non-NSL.

2.1   Consistent vs partial NSL types

Below, we contrast the consistent NSLs, European Portuguese and General Spanish, with their partial varieties – Brazilian Portuguese and Dominican Spanish respectively. We adopt Barbosa's (2011) analysis of French as a non-NSL without further comment.

2.1.1   *Portuguese*

Brazilian Portuguese is classified as a partial NSL, as it only permits NSs in the first person (Holmberg 2005; Kato 1999).

(1) a. *(Eu) acert-ei   a   bola.*
          I   hit-1SG.PST the ball
          "I hit the ball."
     b. *(Você) acert-ou   a   bola.
          you   hit-2SG.PST the ball
          "You (sg) hit the ball."
     c. *(Ele/Ela) acert-ou   a   bola.
          he/she   hit-3SG.PST the ball
          "He/She hit the ball."
     d. *(Nós) acert-amos a   bola.*[4]
          we   hit-1PL.PST the ball
          "We hit the ball."
     e. *(Você-s) acert-aram a   bola.
          you-PL   hit-2PL.PST the ball
          "You (pl) hit the ball."
     f. *(Eles/ Elas)   acert-aram a   bola.[5]
          (they.M/they.F) hit-3PL.PST the ball
          "They (M/F) hit the ball."   [BP]

Note that spoken BP is a partial-NSL closer to the non-NSL end of the spectrum, as it really only permits NSs in the first person singular. This contrasts with European Portuguese, which is a consistent NSL (Barbosa, Duarte & Kato 2005).

---

4.   The pronoun *nós* is not very common; usually *a gente* is used, which cannot be dropped:

   (i) *(A gente) acert-ou   a   bola.
          the person (=we) hit-3SG.PST the ball
          "We hit the ball." [*lit.* "The person hit the ball."]   [BP]

5.   There are some very specific conditions when the third person plural pronouns can be dropped, but in their most common reading, they cannot be omitted.

2.1.2   *Spanish*

With regards to NSs, there is a clear distinction between European and South American Spanish on the one hand, and Caribbean Spanish on the other. In European and South American varieties, overt subjects only occur 19%-27% of the time, while in Caribbean varieties this goes up to 33%–70% (Cabrera-Puche 2008; Mayol 2012; Camacho 2017). Dominican Spanish in particular has very high overt subject usage rates:

Table 3. Average rate of overt and null subject usage in the Dominican Republic (adapted from Camacho 2017)

	El Cibao	Santo Domingo	Average
Overt subjects	70%	68%	69%
Null Subjects	30%	32%	31%

NSs in each of these varieties of Spanish occur for every person/number combination (Toribio 2000). Thus we classify European and South American varieties as consistent NSLs (under the umbrella term 'General Spanish', following Kato 2013), while the Caribbean varieties are classified as partial NSLs.

2.2   EPP type

2.2.1   *French*

We can conclude that French has a DP EPP because there is not V-to-T raising in clauses of all types, because agreement is impoverished, and because French makes use of an expletive. First, although French does have verb raising, it does not occur in non-finite clauses, as shown in (2).

(2)   *Ne   pas posséder   de   voiture…*
      NEG not own.NFIN of.the car
      "Not owning a car…"   [French; Pollock 1989]

Secondly, agreement in French is impoverished (Roberts 2010), indicating that the verb does not have the features needed to check the EPP. We can therefore conclude that V-to-T movement is independent of the EPP (Biberauer & Roberts 2008). Finally, French makes use of an expletive, *il*, indicating it has a DP EPP.[6]

---

6.   See also Davies & Dubinsky (2001), who show that French, alongside English, has the properties of what they call a D-prominent language. D-prominent languages have an EPP checked by a nominal, in contrast to V-prominent languages like Bulgarian and Malagasy.

(3) *Il est   arrivé trois filles.*
     it be.PRES.3SG arrived three girls
     "There have arrived three girls."   [French; Burzio 1986]

2.2.2   *General Spanish*

In this section, we show that at least some varieties of Spanish have a DP-EPP. This is in contrast to Alexiadou & Anagnostopoulou (1998)'s analysis, who argue that consistent NSLs like Spanish have a D-on-V EPP, and that preverbal subjects in these languages do not raise for the EPP, but rather are *always* in an A' position. Of course, preverbal subjects *may* be in some A' position sometimes, but it does not appear to *always* be the case for Spanish (Suñer 2003); much research has shown that there are differences between preverbal subjects and topicalized DPs in Spanish (Goodall 2001; Camacho 2006, *inter alia*).

One of Alexiadou & Anagnostopoulou's (1998) pieces of evidence that preverbal subjects are topicalized is that adverbs can intervene between the subject and the verb. However, in Spanish, negative quantifier subjects (which cannot be topicalized) must be adjacent to the verb, as in (4).

(4) *Nadie*   (*casi) *pudo avanzar 3 metros.*
     No-one almost could advance 3 metres
     "No one could advance 3 metres."   [Spanish; Camacho 2006]

Thus, preverbal subjects in Spanish appear to be in A-position at least some of the time. Since Alexiadou & Anagnostopoulou's (1998) analysis required all preverbal subjects to be in A'-positions, their analysis of Spanish is untenable. Furthermore, Spanish patterns with other DP-EPP languages in the behaviour of non-nominal subjects.

Non-nominal subjects in Spanish have parallel properties to those in French and English, indicating that they are able to check the EPP. For example, some locatives and other adverbials can occupy the preverbal position and exhibit nominal properties such as plural agreement when they do, as in (5).

(5) *Bajo   la   cama y   detrás del   ropero   todavía* **están**
     under the.F bed   and behind of.the.M wardrobe still   **be.3PL**
     por limpiar.
     for clean.NFIN
     "Under the bed and behind the wardrobe still need to be cleaned."
                                    [Spanish (Rioplatense); María C. Cuervo, p.c.]

Such non-nominal subjects are also able to undergo subject-to-subject raising, indicating that they are in an A-position.

(6) *Bajo   la   cama y   en el   ropero   **parecen**   ser*
    under the.F bed   and in the.M wardrobe **seem.PRES.3PL** be.NFIN
    buenos   lugares para esconderse.
    good.M.PL place.PL for   hide.NFIN-REFL
    "Under the bed and in the wardrobe seem to be good places to hide."
                           [Spanish (Rioplatense); María C. Cuervo, p.c.]

Spanish also shares with Finnish the ability to have referential adverbials in subject position, which Holmberg (2005) argues check the EPP in Finnish. Goodall (2001) and Sheehan (2007) both argue that null versions of these locatives check the EPP in Spanish when no overt DP is present.

   Note there is variation among dialects (e.g., Example (6) was judged ungrammatical by a speaker of Mexican Spanish); however, at least *some* dialects of Spanish have a DP EPP. In fact, that the EPP type varies across dialects in which the NS status stays constant is further evidence that NSL type and EPP type are not co-extensive.

2.2.3   *Brazilian Portuguese*

The EPP works differently in BP due to its topic-prominent nature, although the EPP type is still DP-EPP. The analysis of BP presented here has two ingredients: (a) the optional deletion of the specifier of the root node and (b) the ability of topics to check the EPP.

   Sheehan (2007) argues that the specifier of the root node can be deleted, similar to 'diary drop' in English (Haegeman 2000), and languages such as German, Swedish, and Dutch (Ross 1982). This results in sentences with NSs such as the following:

(7) *(Eu) comprei   os livros ontem*
    I   bought.1SG the books yesterday
    "I bought the books yesterday."   [BP; Sheehan 2007: 260]

This also allows for NSs in embedded clauses, as in (8b), if there is a salient linguistic antecedent (8a). The embedded subject can then be a topic that moves to the matrix topic position (cf. Rodrigues 2004), which is then deleted as specifier of the root node.

(8) a. *E   o   Paulo₁?*
         And the Paulo
         "What about Paulo?"
    b. *A   Maria₂ disse que EC₁ estava doente.*
         the Maria said that   was   ill
         "Maria said he was ill."   [BP; Sheehan 2007: 265]

The topic may also remain overt, as shown in (9).

(9) *A   Maria$_1$, o   José disse que EC$_1$ comprou um carro.*
     The Maria, the José said that   bought a   car
     "Maria, José said that she bought a car."   [BP; Sheehan 2007: 266]

NSs in BP behave like bound variables; they require a linguistic (not pragmatic) antecedent and display the properties of obligatory control (Sheehan 2007). There is thus a syntactically active NS present in these constructions.

However, BP does not pattern with other DP-EPP languages such as Spanish, English and French, in the behaviour of non-nominal subjects. For example, (at least some) speakers of BP don't appear to have plural agreement with coordinated non-nominal subjects, unlike French and (some varieties of) Spanish.

(10) *Embaixo da   cama e   atrás do   armário ainda*
      Under   of.the.F bed   and behind of.the.M cabinet still
      precisa   limpar.
      need.3SG clean.NFIN
      "Under the bed and behind the wardrobe still need to be cleaned."

                                                          [BP; Suzi Lima, p.c.]

This is likely due to the topic-prominent nature of BP. Whereas the non-nominal subjects of French, Spanish and English are interpreted as covert nominals, triggering plural agreement, they are interpreted as topics in BP, and agreement is controlled by the postverbal thematic subject. These agreement facts parallel light locative inversion in English, where, according to Bruno (2016), a locative PP raises to a topic position in spec,CP, but passes through spec,TP, checking the EPP. In these cases, agreement is also controlled by the post-verbal thematic subject, as shown in (11), although the locative topic checks the EPP.

(11) a. *Into the room walks Robin.*
      b. *Into the room walk Batman and Robin.*

Additionally, some speakers of BP[7] (in contrast to European Portuguese) can have non-thematic deictic nominals in the subject position in out-of-the-blue contexts, as shown in (12).

(12) *As   florestas chovem muito.*
      the.F.PL forest.F rain.3PL much
      "It rains a lot in the forests."
                              [BP; Costa 2010, as cited in Naves, Pilati & Salles 2013]

---

7. Although there seems to be some understudied variation (Suzi Lima, p.c.).

These deictic topics disrupt the ability of an embedded clause subject to corefer with the matrix clause subject.

(13) a.   *O   Pedro$_i$ disse   que conserta$_{i/*j}$ sapato.*
        the.M Pedro say.PST.3SG that fix.3SG   shoe
        "Pedro$_1$ said that he$_1$ fixes shoes."
    b.   *O   Pedro$_1$ disse   que aqui conserta$_{*1/2}$ sapato.*
        the.M Pedro say.PST.3SG that here fix.3SG   shoe
        "Pedro said that s/he fixes shoes here."   [BP; Naves et al. 2013]

The deictic nominal acts as an intervener; the NS cannot be interpreted as a referential subject which raises to topic position and is deleted. Instead, the deictic nominal checks the EPP.

BP has a DP-EPP, although the EPP effects of BP interact with topicalization processes, giving BP different properties than other DP-EPP languages. In BP, a topic can be null since it raises to a topic position, which may be subsequently deleted, as specifier of the root node. Non-nominal EPP-checkers in BP do not display properties of non-nominal subjects, because they are not subjects, but topics that check the EPP en route to the topicalized position.

3.   Synchronic and diachronic variation

One of our claims is that EPP type and NSL type can vary independently across varieties, as well as over time, providing evidence that these types are not co-extensive.

Consistent NSLs can become a non-NSL or a partial NSL over time, as in French (Kato 1999 and King, Martineau & Mougeon 2011) and BP, respectively (Barbosa et al. 2005; Kato 1999, 2013). This change may be explained in one of two ways, as described in (14).

(14)   Null Subject Chicken-Egg Problem (Bilgin 2017)
    a.   Hypothesis 1 (H1): Lose Person-agreement Markers First
        Person-agreement markers are lost, which can trigger the usage of overt pronouns.
    b.   Hypothesis 2 (H2): Increase Overt Pronoun Usage First
        Overt pronouns become more common, thereby rendering distinct person-agreement markers redundant, which may cause these markers to fall into disuse, and eventually disappear.

In the literature, H1 is easily assumed; it is generally believed that when a language loses its distinct person-agreement markers, it can cause the use of overt pronouns, causing languages that were once consistent NSLs to become non-NSLs or partial

NSLs (Kato 1999; Rodrigues 2004; Barbosa et al. 2005; Barbosa 2009; Holmberg et al. 2009; King et al. 2011, *inter alia*). We claim that H2 is also possible, demonstrated through the Goiás dialect of Brazilian Portuguese and Dominican Spanish. In both varieties, the NSL type changes, but the EPP does not (right away), providing evidence that they are not co-extensive.

3.1   Diachronic variation

3.1.1   *Goiás BP*
Borges & Pires (2017) show that, in the Goiás dialect of Brazilian Portuguese (BP), an increase in the occurrence of overt subjects and the loss of free inversion (VS orders) *preceded* the impoverishment of the verbal paradigm. Although cases of VS order in the 19th century did exist, they mostly involved locative inversion, as shown in (15). In these cases, the locative may be checking the EPP.

(15)   *Dia 22 chegou o   presidente e   mais alguns colegas.*
       day 22 arrived the president and more some   colleagues
       "On the 22nd, the president and some (of his) colleagues arrived."

                                    [Goiás BP (19th C); Borges & Pires 2017: 10]

However, Borges & Pires (2017) note that there was no lack of agreement in their 18th century data, and only marginal impoverishment in the 19th century, with only 15% (54/380) of clauses with an overt plural subject lacking agreement. Furthermore, all of these involved third person unaccusative or existential/locative verbs.

Borges & Pires (2017) propose that the change from consistent to partial NSL in Goiás BP was triggered by the loss of a [D] feature in T. Since this [D] feature was no longer available for the EPP, it caused a rise in the prevalence of overt preverbal subjects. This change was *not* triggered by loss of agreement, as agreement was not yet lost. Furthermore, Borges & Pires (2017) show that these changes were also accompanied by the marginalization of overt strategies for generic subjects (*se* constructions), another characteristic of consistent NSLs. These were replaced with an innovative impersonal construction with a NS. We therefore propose that the loss of the [D] feature in T is a change in EPP type. Crucially, then, the change in EPP type preceded the change in agreement in Goiás. The change in EPP type led to a decrease in NSs, which, in turn, triggered a change in NSL type, and finally, the loss of rich agreement.

3.1.2   *Dominican Spanish*

In this section, we show that the NSL type of Dominican Spanish changed without changing its EPP type first, following H2 in (14). Toribio (2000) shows that although Dominican Spanish has lost most of its distinct agreement morphology, NSs are still allowed (Table 4). Note that the brackets indicate the availability of NSs paired with their respective agreement suffixes:

Table 4. Agreement in General Spanish (excluding Caribbean Spanish, see § 2.1.2) and Dominican Spanish (based on Kattan-Ibarra & Pountain 2003: 34, 59; Pöll 2015: 331; and Toribio 2000: 317–322, from Bilgin 2017)

	General Spanish		Dominican Spanish	
1SG	*(yo)*	*-o*	*(yo)*	*-o*
2SG	*(tú)*	*-as*	*(tú)*	*-a*
3SG	*(él/ella)*	*-a*	*(él/ella)*	*-a*
1PL	*(nosotros/-as)*	*-amos*	*(nosotros/-as)*	*-amo*
2PL	*(vosotros/-as)*	*-áis*	*(ustedes)*	*-a(n)*
	(ustedes)	*-an*		
3PL	*(ellos/ellas)*	*-an*	*(ellos/ellas)*	*-a(n)*

Toribio (2000) notes that, even where distinct agreement morphology is lost, Dominican Spanish subjects may be null. If distinct person-agreement morphology licenses NSs, as H1 suggests, then we would expect the second and third person pronouns to always be overt since they lack distinct agreement morphology. Rather, Dominican Spanish has followed the path we sketched out in H2, in which overt pronouns gain wider usage first, which in turn is followed by the decline of agreement morphology (for further evidence, see Bilgin 2017: 87–100; Camacho 2017).[8]

Recall that the EPP type of General Spanish is DP EPP (§ 2.2.2). Like General Spanish, the EPP type of Dominican Spanish would have been DP EPP in the past. We argue that the EPP type of Dominican Spanish has remained constant since its agreement is impoverished and cannot satisfy the EPP. Meanwhile, it has clearly developed into a partial NSL from a consistent NSL. Crucially, this analysis demonstrates that NSL type may change without a change in EPP type first.[9]

---

8. For a pragmatics and discourse-based analysis of the change in Dominican Spanish, see Corr (2015). Unfortunately, we cannot discuss this analysis further due to space limitations.

9. The trigger may be as suggested by Mayol (2012), who argues that the use of overt subjects increased in Caribbean Spanish and Brazilian Portuguese through contact with non-NSL African languages.

3.2   Synchronic variation

In this section, we show that different NSL and EPP types co-exist within the same variety and even the same speaker in Dominican Spanish. This shows that a change in NSL type does not necessarily have an immediate impact on EPP type, and vice versa.

Toribio (2000) observes that Dominican Spanish sometimes behaves like a non-NSL and sometimes like an NSL, synchronically. For example, the utterances in (16) exhibit variation, although they are all produced by the same speaker.

(16) Null Subjects
  a. *… vino con veintiocho mil dólares …*
      " … he came with twenty-eight dollars … "
  b. *… no se le puede echar plata …*
      " … you can't invest money … "
      Overt Subjects:
  c. *Cuando él vino la primera vez …*
      "When he came for the first time … "
  d. *… y yo le dije que …*
      " … and I told him that … " [Spanish (Dominican); Toribio 2000: 337–8]

There are even instances of overt expletive subjects in Dominican Spanish (17) and non-finite clauses with overt pre-verbal nominative subjects (18) (Toribio 2000), both of which are not usually found in other varieties of Spanish.

(17) a. *Ello llegan   guaguas hasta allá.*
          EXPL arrive.3PL bus.PL   until   there
                                  [Spanish (Dominican); Toribio 2000: 321]
  b. cf. *Llegan   guaguas hastá [sic] allá.*
          arrive.3PL bus.PL   until   there
          "There arrive busses there."   [Spanish; Toribio 2000: 321]

(18) a. *Ven   acá, para nosotros ver-te.*
          come.IMP.2SG here for   1PL.NOM see.NFIN-2SG
                                  [Spanish (Dominican); Toribio 2000: 323]
  b. cf. *Ven   acá, para ver-te   (nosotros).*
          come.IMP.2SG here for   see.NFIN-2SG 1PL.NOM
          "Come here, for us to see you."   [Spanish; Toribio 2000: 323]

Based on such examples, Toribio (2000) argues that Dominican speakers have two distinct I-languages, similar to bilingualism, which she terms being 'intralingual' or 'intra-dialectal'. This free alternation between different NSL types within Dominican Spanish suggests that various NSL types can exist at the same time in the same language.

Crucially, agreement in Dominican Spanish is impoverished, whether the subject is null or overt; thus, the EPP type remains constant, while the NSL type varies. This, again, indicates that strict co-extensiveness between NSL and EPP types is not empirically motivated.

4. Conclusion

In this chapter, we have simply shown that NSL type can vary while EPP type remains constant. There is one obvious alternative analysis of these facts: that D-on-V EPP doesn't exist. But we do not think this is the case. First of all, the D-on-V EPP analysis was developed predominantly for Greek, and then was extended to Romance, somewhat on the assumption that consistent NSLs are the same. A more thorough analysis of Greek is needed before further conclusions are drawn, especially since the status of the EPP in Greek is also contested (Spyropoulos & Philippaki-Warburton 2001).[10] Secondly, Richards & Biberauer (2005) argue for a typology of EPP-checking mechanisms in Germanic which includes D-on-V EPP with pied-piping for languages such as Icelandic. They arrive at this conclusion by comparing Germanic languages which raise the vP to spec,TP; some require an overt DP in spec,vP, while others require rich inflection. Finally, there are cases of synchronic and diachronic variation (discussed here, but see also Richards & Biberauer 2005), where EPP type appears to change, implying that another type of EPP exists. Furthermore, another option not addressed in this chapter is that a language may allow more than one mode of EPP-checking (Sheehan 2016; Doner 2019).

Further research is needed in order to determine if the relation between NSL type and EPP type is a double dissociation, and if all possible combinations exist. We predict the existence of languages with the same NSL type, but different EPP types. For example, Greek and Spanish are both consistent NSLs, although Greek may have D-on-V EPP (Alexiadou & Anagnostopoulou 1998), while Spanish has DP EPP (Goodall 2001). Once this typology has been completed, any patterns and gaps that remain will need to be explained. However, it cannot be assumed that EPP type and NSL type are correlated unless evidence is provided. Correlation should *not* be the null hypothesis.

---

10. Spyropoulos & Philippaki-Warburton (2001) argue that all preverbal subjects in Greek occupy topic positions, as would be expected in a V-on-D EPP analysis of Greek. Their arguments for a DP EPP assume the Lexicalist hypothesis that word-internal morphemes are unable to participate in syntactic operations and that EPP-checking elements must be referential, neither of which are assumptions in Doner's (2019) typology of the EPP.

Acknowledgements

Abbreviations

1,2,3	first, second, third person	M	masculine
IMP	imperative mood	PRES	present tense
NEG	negation	PL	plural
NOM	nominative case	PST	past tense
NFIN	non-finite	F	feminine
REFL	reflexive	FUT	future tense
DAT	dative case	SG	singular
EXPL	expletive	SUBJ	subjunctive.

References

Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parametrizing agr: Word order, v-movement and EPP-checking. *Natural Language and Linguistic Theory* 16(3). 491–539. https://doi.org/10.1023/A:1006090432389

Barbosa, Pilar. 2009. Two kinds of subject pro. *Studia Linguistica* 63(1). 2–58. https://doi.org/10.1111/j.1467-9582.2008.01153.x

Barbosa, Pilar. 2011. Pro-drop and theories of pro in the Minimalist Program Part 1: Consistent null subject languages and the pronominal-agr hypothesis. *Language and Linguistics* 5(8). 551–570.

Barbosa, Pilar, Maria Duarte & Mary Aizawa Kato. 2005. Null subjects in European and Brazilian Portuguese. *Journal of Portuguese Linguistics* 4. 11–52. https://doi.org/10.5334/jpl.158

Biberauer, Theresa. 2010. Semi null-subject languages, expletives, and expletive *pro* reconsidered. In Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.), *Parametric variation: Null Subjects in Minimalist Theory*, 153–199. Cambridge: Cambridge University Press.

Biberauer, Theresa & Ian Roberts. 2008. Subject, tense and verb movement in Germanic and Romance. *Cambridge Occasional Papers in Linguistics* 3. 24–43.

Bilgin, Çağrı. 2017. *Synchro-diachronic approach to null subjects: An account of null subject language diversity and associated diachronic change*. Toronto: University of Toronto MA forum paper.

Borges, Humberto & Acrisio Pires. 2017. The emergence of Brazilian Portuguese: Earlier evidence for the development of a partial null subject grammar. *Proceedings of the Linguistic Society of America* 2(31). 1–15. https://doi.org/10.3765/plsa.v2i0.4096

Bruno, Chris. 2016. Three kinds of locative preposing. *Toronto Working Papers in Linguistics* 36.

Burzio, Luigi. 1986. *Italian syntax*. Reidel: Dordrecht. https://doi.org/10.1007/978-94-009-4522-7

Cabrera-Puche, Maria. 2008. Null subject patterns in language contact: The case of Dominican Spanish. New Brunswick, NJ: Rutgers University dissertation.

Camacho, José. 2006. Do subjects have a place in Spanish? In Chiyo Nishida & Jean-Pierre Y. Montreuil (eds.), *New Perspectives on Romance Linguistics: Volume I: Morphology, syntax, semantics, and pragmatics. Selected papers from the 35th Linguistic Symposium on Romance Languages (LSRL). Austin, Texas, February 2005*, 51–66. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.275.06cam

Camacho, José. 2017. The null subject parameter revisited: The evolution from null subject Spanish and Portuguese to Dominican Spanish and Brazilian Portuguese. In Mary Aizawa Kato & Francisco Ordóñez, (eds.) *The morphosyntax of Portuguese and Spanish in Latin America*, 27–48. New York: Oxford University Press.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.

Corr, Alice. 2015. Overt expletives in Ibero-Romance: A diachronic and diatopic perspective. *Revue Roumaine de Linguistique* 60 (2–3). 205–222.

Costa, João. 2010. PB e PE: orientação para o discurso importa? [BP and EP: Does discourse orientation matter?] *Estudos de Línga(gem)* 8(1). 123–143. https://doi.org/10.22481/el.v8i1.1117

Davies, William & Stanley Dubinsky. 2001. Functional architecture and the distribution of subject properties. In William Davies & Stanley Dubinsky (eds.). *Objects and other subjects: Grammatical functions, functional categories, and configurationality*, 247–279. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-010-0991-1_10

Doner, Julianne. 2019. The EPP across languages. Toronto: University of Toronto dissertation.

Goodall, Grant. 2001. The EPP in Spanish. In William Davies & Stanley Dubinsky (eds.). *Objects and other subjects: Grammatical functions, functional categories, and configurationality*, 193–223. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-010-0991-1_8

Haegeman, Liliane. 2000. Adult null subjects in non pro-drop languages. In Marc-Ariel Friedemann & Luigi Rizzi, *The acquisition of syntax: Studies in comparative developmental linguistics*, 129–169. Harlow: Longman.

Haegeman, Liliane. 2013. The syntax of registers: Diary subject omission and the privilege of the root. *Lingua* 130. 88–110. https://doi.org/10.1016/j.lingua.2013.01.005

Holmberg, Anders. 2005. Is there a little *pro*? Evidence from Finnish. *Linguistic Inquiry* 36. 533–564. https://doi.org/10.1162/002438905774464322

Holmberg, Anders, Aarti Nayudu & Michelle Sheehan. 2009. Three partial null-subject languages: A comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica* 63. 59–97. https://doi.org/10.1111/j.1467-9582.2008.01154.x

Kato, Mary Aizawa. 1999. Strong pronouns, weak pronominals and the null subject parameter. *Probus* 11. 1–37. https://doi.org/10.1515/prbs.1999.11.1.1

Kato, Mary Aizawa. 2013. Brazilian Portuguese and Caribbean Spanish: Similar changes in Romania Nova. In Charlotte Galves, Sonia Cyrino, Ruth Lopes, Filomena Sandalo & Juanito Avelar (eds.), *Parameter theory and linguistic change*, 117–132. Oxford: Oxford University Press.

Kattan-Ibarra, Juan & Christopher Pountain. 2003. *Modern Spanish grammar: A practical guide*. 2nd edn. New York: Routledge. https://doi.org/10.4324/9780203137352

King, Ruth, France Martineau & Raymond Mougeon. 2011. The interplay of internal and external factors in grammatical change: First person plural pronouns in French. *Language* 87(3). 470–509. https://doi.org/10.1353/lan.2011.0072

Massam, Diane & Carolyn Smallwood. 1997. Essential features of predication in English and Niuean. In Kiyomi Kusumoto (ed.), *Proceedings of NELS* 27, 263–272. Amherst: Graduate Linguistics Student Union, University of Massachusetts.

Mayol, Laia. 2012. An account of variation in the rates of overt subject pronouns in Romance. *Spanish in Context* 9(3). 420–442. https://doi.org/10.1075/sic.9.3.03may

Naves, Rozana, Eloisa Pilati & Heloise Salles. 2013. 'As cidades da Amazônia chovem muito': sujeitos não temáticos e as propriedades de INFL no Português do Brasil. ['The cities of the Amazon rain a lot': Non-thematic subjects and the properties of INFL in Brazilian Portuguese.] Abstract. Paper presented at the Portuguese Linguistics in the United States (PLUS) Conference, Athens, GA, November 14–16. https://portugueselinguistics.files.wordpress.com/2013/11/plus-program-abstracts-nov7.pdf. Accessed April 23, 2018.

Pöll, Bernhard. 2015. Caribbean Spanish = Brazilian Portuguese? Some comparative thoughts on the loss of pro-drop. *Studies in Hispanic and Lusophone Linguistics* 8(2). 317–354. https://doi.org/10.1515/shll-2015-0012

Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 20. 365–424.

Richards, Marc & Theresa Biberauer. 2005. Explaining *expl*. In Marcel den Dikken & Christina Tortora (eds.), *The function of function words and functional categories*, 115–154. Amsterdam: John Benjamins. https://doi.org/10.1075/la.78.06ric

Roberts, Ian. 2010. Varieties of French and the null subject parameter. In Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.), *Parametric variation: Null subjects in minimalist theory*, 303–327. Cambridge: Cambridge University Press.

Rodrigues, Cilene. 2004. Impoverished morphology and a-movement out of case domains. College Park, MD: University of Maryland dissertation.

Ross, John. 1982. Pronoun deleting processes in German. Paper presented at the *Annual Meeting of the Linguistic Society of America*, San Diego, CA, December 27–30.

Sheehan, Michelle. 2007. The EPP and null subjects in romance. Newcastle upon Tyne, UK: Newcastle University dissertation.

Sheehan, Michelle. 2016. Subjects, null-subjects and expletives in Romance. In Susann Fischer & Cristoph Gabriel (eds.). *Manual of Grammatical Interfaces in Romance* (*Manuals of Romance Linguistics* 10), 329–362. Berlin: De Gruyter. https://doi.org/10.1515/9783110311860-014

Spyropoulos, Vassilios & Irene Philippaki-Warburton. 2001. 'Subject' and EPP in Greek: The discontinuous subject hypothesis. *Journal of Greek Linguistics* 2. 149–186.

Suñer, Margarita. 2003. The lexical preverbal subject in a Romance null subject language: Where are thou? In Rafael Núñez-Cedeño, Luis López & Richard Cameron (eds.), *A Romance perspective on language knowledge and use: Selected papers from the 31st Linguistic Symposium on Romance Languages (LSRL), Chicago, 19–22 April 2001*, 341–357. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.238.25sun

Toribio, Almeida Jacqueline. 2000. Setting parametric limits on dialectal variation in Spanish. *Lingua* 110(5). 315–341. https://doi.org/10.1016/S0024-3841(99)00044-3

Weir, Andrew. 2012. Left-edge deletion in English and subject omission in diaries. *English Language and Linguistics* 16(1). 105–129. https://doi.org/10.1017/S136067431100030X

On (un)grammatical sequences of *se*s in Spanish

Jonathan E. MacDonald and Almike Vázquez-Lozares

University of Illinois at Urbana-Champaign

We discuss the ungrammaticality of Spanish impersonal *se* (Imp$_{se}$) in combination with another *se* in control structures. Our main goal is to critically evaluate Martins & Nunes's (2017) account of parallel phenomena in European Portuguese, who, adopting a movement theory of control, claim that Imp$_{se}$ merges in the external argument of the non-finite embedded verb and moves to the matrix clause. When Imp$_{se}$ and the other *se* are transferred within the same phase, the result is ungrammaticality due to an identity violation. We illustrate that their approach cannot be extended to Spanish. Instead, we put forth MacDonald & Vázquez-Lozares's (2021) proposal as an alternative, which as we show allows for an extension to ungrammatical contexts in which the identity of *se* cannot be appealed to.

Keywords: impersonal *se*, paradigmatic *se*, clitic climbing, clitic combinations, identity violation, control structures, Spanish

1. Introduction

The main set of phenomena that we discuss in this paper is illustrated in (1), from Spanish.[1]

(1) a. *Se   se   levanta pronto en este país.*
     Imp$_{se}$ Refl$_{se}$ raises early in this country
     Intended: "One gets up early in this country."

---

1. Here and throughout, the data presented as Spanish without specific reference to a particular variety comes from standard peninsular Spanish, which, at least with respect to the data presented, is indicative of general patterns not specific to peninsular Spanish. Where there is variation, or the data do not come from standard peninsular Spanish, we specify which variety is under discussion, as we do in § 1.1.2.

      b. *Se   *intentó sentarse en un buen sitio.*
           Imp$_{se}$ tried   sit.Refl$_{se}$ in a   good spot
           Intended: "One tried to sit in a good spot."

We see in (1a) that a sequence of impersonal *se* (Imp$_{se}$) plus an (inherently) reflexive *se* (Refl$_{se}$) clitic is ungrammatical.[2] We see in (1b) that when a control verb is impersonalized with Imp$_{se}$, the embedded verb cannot have Refl$_{se}$ either; there is an effect 'at a distance'.

    In a recent proposal, Martins & Nunes (2017), henceforth M&N, argue that the patterns in (1) result from an identity violation in the presence of identical clitics when transferred within the same strong phase, where *v*P and CP are the relevant phases.[3] We offer arguments against extending their approach to Spanish. We note that an alternative proposal from MacDonald & Vázquez-Lozares (2021), where it is argued that Imp$_{se}$ lacks number, can explain why Refl$_{se}$ (and any other paradigmatic *se*) is ungrammatical with Imp$_{se}$ – because its number feature remains unvalued.[4] Moreover, as we note, this approach also offers an explanation for the ungrammaticality of non-clitic reflexive *sí* with Imp$_{se}$ in Spanish, as in (2). No appeal to identity avoidance can be made, since there are no identical clitics, yet we see effects 'at a distance', just like we see in (1b) above.

(2) a. *Se   *habla de sí mismo.*
         Imp$_{se}$ speaks of self same.
         "One speaks about oneself."
    b. *Se   *intenta hablar de sí   mismo.*
         Imp$_{se}$ tries   speak of self same
         "One tries to speak about oneself."

---

2. What we call impersonal *se* Cinque (1988) and Dobrovie-Sorin (1998) call [-arg] *si* and Nominative *si* respectively. Martins & Nunes (2017) refer to it as indefinite *se*.

3. In M&N's account, the identity of clitics depends on both morphological and phonological information. We assume the same, although for the majority of the cases of Imp$_{se}$ we discuss here, identity in phonological form suffices. Importantly, however, M&N assume that morphological information plays a role in the calculation of identity. This justifies looking at other morphological effects in the presence of Imp$_{se}$ plus non-identical clitics, something we do in § 1.1.2 below.

4. MacDonald & Vázquez-Lozares (2021) assume that it is not *se* itself that lacks number, but a corresponding null implicit projected *pro*. See also Mendikoetxea (2008); MacDonald (2017); MacDonald & Maddox (2018); Ordóñez (2021) for the assumption that there is a *pro* in these constructions. Nevertheless, note that the patterns discussed here can also be explained by assuming that *se* itself lacks number. As far as we can tell the data discussed in the present paper do not distinguish between the two approaches.

The structure of the paper is the following. In § 1 we lay out our arguments for not extending M&N's account to Spanish. We discuss predictions generated by their claim that Imp$_{se}$ merges in the external argument position of the embedded clause. We look specifically at sequences of Imp$_{se}$ with so-called spurious *se* and accusative direct object clitic *lo*. The predictions are not born out. We conclude that we should not take Imp$_{se}$ to be merged in the embedded clause. In § 2, we discuss clitic climbing contexts, where we see that Imp$_{se}$ does not pattern with other climbing clitics, something we suggest might be expected on M&N's approach. In § 3, we indicate an alternative analysis and discuss an extension that does not appeal to the identical status of clitics. Section 4 briefly recaps.

2.   Martins & Nunes (2017)

According to M&N, the patterns in (1) result from an identity violation caused by identical clitics being transferred within the same strong phase, where *v*P and CP are the relevant phases. An initial expectation of this proposal is that when two *se*s are in different clauses, as in (3), no identity violation arises.[5]

(3) *Se   dijo que se   sentó en el   suelo.*
     Imp$_{se}$ said that Refl$_{se}$ sat   in the floor.
     "It was said that s/he sat on the floor."

That the clitics in (1a) are transferred within the same phase is obvious, but less so for the clitics in (1b). Assuming the movement theory of control, M&N claim that Imp$_{se}$ originates in Spec,v of the embedded clause, as illustrated in (4a), and then moves to the matrix clause. (4b) illustrates Transfer of the embedded clause, which is parallel to the matrix context in (4c).

(4) a.   [$_{vP}$ v [$_{VP}$ intentó [$_{CP}$ C [$_{TP}$ **Imp$_{se}$ T-Refl$_{se}$** [$_{vP}$ <Imp$_{se}$> v [$_{VP}$ sentar-<Refl$_{se}$>]]]]]]
    b.   Transfer of Embedded TP: [$_{TP}$ **Imp$_{se}$ T-Refl$_{se}$** …]
    c.   Transfer of Matrix CP: [$_{CP}$ C [$_{TP}$ **Imp$_{se}$-Refl$_{se}$-T**…]]

Imp$_{se}$ and Refl$_{se}$ are transferred together within the same phase, which leads to identity violations.

---

5.   Consider the Vicentino data in (i) from Pescarini (2011: 12), where the sequence *se se* is grammatical. The data raise questions about whether in all languages the phase should be taken as the domain for analyzing patterns parallel to (1).

(i) *Se   se   lava le man.*
     himself.CL one.CL wash the hands
     "You(imp.) wash your hands."

We focus on two crucial aspects of M&N's account in this paper. The first is that the observed effects of two *se* in matrix contexts will also be observed and be the same in embedded contexts, something that we refer to as a matrix-embedded symmetry. The second is that Imp_{se} originates in the embedded clause and moves to the matrix clause. Based on these two aspects of their proposal, two expectations are generated: (1) since morphological information plays a role in the calculation of identity (see footnote 2), the matrix-embedded symmetry observed for Imp_{se} plus identical clitics should extend to other morphological effects observed for Imp_{se} plus any other clitic, whether identical with Imp_{se} or not; and (2) since Imp_{se} moves from the embedded clause to the matrix clause, we might expect it to pattern with other climbing clitics.

2.1   Matrix-embedded asymmetries

In this section, we illustrate two cases where we see a matrix-embedded asymmetry between a sequence of Imp_{se} plus clitic. In § 2.1.1, we look at sequences of Imp_{se} plus spurious *se* ($Spur_{se}$). In § 2.1.2, we look at sequences of Imp_{se} plus accusative clitic *lo*.

2.1.1 Imp_{se} + $Spur_{se}$ matrix-embedded asymmetry
Spanish $Spur_{se}$ surfaces when there are two 3rd person clitics, as illustrated in (5) (Perlmutter 1971; Bonet 1991; Nevins 2007; Walkow 2012; Pancheva & Zubizarreta 2017, *inter alia*.)

(5) a. *Mandamos un regalo   a   Juan.*
         Sent   a gift   to Juan.
         "We sent a gift to Juan."
     b. **Le   lo mandamos.* → c. *Se   lo mandamos.*
         Him   it sent $Spur_{se}$ it sent
         "We sent him it."   "We sent him it."

In (5b), both the direct and indirect objects from (5a) are pronominalized, resulting in the ungrammatical sequence of two third person clitics. To 'save' the derivation, $Spur_{se}$ surfaces in place of the indirect object clitic *le*, as illustrated in (5c). Observe that when (5c) is impersonalized, as in (6) below, the result is a sequence of Imp_{se} plus $Spur_{se}$. This leads to ungrammaticality. Moreover, it looks like an identity violation of the type discussed in M&N.[6]

---

6. According to M&N (2017: 640), " … identity avoidance involving indefinite se [=Imp_{se}] does not trigger deletion as a rescue strategy and its effects are identified simply via … unacceptability … "; Nevins (2007) argues extensively that the features causing the problem giving

(6) **Se   se   lo manda todos los años.*
    Imp$_{se}$ Spur$_{se}$ it sends all   the years.
    "They send it to him every year."

In fact, M&N can easily handle this sequence of identical clitics in the matrix clause, since both will be transferred within the matrix CP phase, as illustrated in (7).

(7) a. [$_{CP}$ C [$_{TP}$ **Imp$_{se}$-Spur$_{se}$-lo-T** [$_{vP}$ <**Imp$_{se}$**> v [$_{VP}$ manda-<le><lo>…]]]]]]
   b. Transfer of the matrix CP:
      [$_{CP}$ **Imp$_{se}$-Spur$_{se}$-lo-T** [$_{vP}$ <**Imp$_{se}$**>]…]

Importantly, observe that in (8) the sequence of Imp$_{se}$ plus *le* is grammatical, which would force one to conclude that the calculation of the identity violation must take place after the Spur$_{se}$ rule changes *le* to *se*, otherwise, (6) would be grammatical.

(8) *Se   le   manda un regalo todos los años.*
    Imp$_{se}$ him send   a   gift   all   the years
    "They send him a gift every year."

Thus, in terms of the ordering of the rules to account for the ungrammatical sequence of Imp$_{se}$ plus Spur$_{se}$ in matrix context, M&N are forced to conclude the order in (9).

(9) **Matrix Rule Order**
    1st: Apply Spurious *se* rule
    2nd: Calculate identity violation

Now consider the data in (10), where we see no Imp$_{se}$ plus Spur$_{se}$ identity violation.

(10) *Se   intentó mandárselo.*
    Imp$_{se}$ tried   send.Spur$_{se}$.it
    "They tried to send it to him."

Consider the portion of the embedded clause in (11) which transfers when matrix v merges.

(11) a. [$_{vP}$ v [$_{VP}$ intentó [$_{CP}$ C [$_{TP}$ **Imp$_{se}$ T-le-lo** [$_{vP}$ <**Imp$_{se}$**> v [$_{VP}$ mandar-<le><lo>]]]]]]]
   b. Transfer of embedded TP:
      [$_{TP}$ **Imp$_{se}$ T-le-lo** [$_{vP}$ <**Imp$_{se}$**>]…]

---

rise to Spur$_{se}$ are morphosyntactic features relating to 3rd person. He claims that the repair, i.e., the morphological rule, erases *le*'s person features changing it to Imp$_{se}$. Imp$_{se}$ and Spur$_{se}$ are morphosyntactically identical on his account. For a different approach, see Cuervo (2013).

On M&N's account (11b) must be what is transferred in the embedded TP phase, since the sentence in (10) is grammatical and we know that the sequence of Imp_{se} plus $Spur_{se}$ is ungrammatical. That is, the calculation of identity must have taken place before the $Spur_{se}$ rule applies in these embedded contexts. This leads to the embedded rule ordering in (12a). By comparing it to the matrix rule ordering in (12b), it becomes obvious that there is a conflict.

(12)   a.   **Embedded rule order**
             1st: Calculate identity violation
             2nd: Apply Spurious *se* rule
       b.   **Matrix rule order**
             1st: Apply Spurious *se* rule
             2nd: Calculate identity violation

Contradictory rule orderings are necessary to account for the Imp_{se} plus $Spur_{se}$ clitic sequence facts in matrix and embedded clauses on M&N's account. This rule ordering paradox results from the Imp_{se} plus $Spur_{se}$ matrix-embedded asymmetry, not expected on their account.

2.1.2   *Imp_{se} + lo matrix-embedded asymmetry*

Consider a contrast between Honduran Spanish (Hond) and Rioplatense Spanish (Riopl) in (13) with respect to Imp_{se} plus direct object *lo*.[7]

(13)   a.   %*Se lo quiere abrazar.*   [√Riopl/*Hond]
       b.   *Se   quiere abrazarlo.*
             Imp_{se} wants hug.him
             "One wants to hug him."

In both varieties, (13b) is grammatical, when the clitic remains in the embedded clause. The contrast arises when the clitic moves into the same domain as Imp_{se}. It is grammatical in Rioplatense, but ungrammatical in Honduran Spanish, as illustrated in (13a).[8]

Focusing on the Honduran Spanish patterns for now, compare what the transfer of matrix and embedded clauses in (14a) and (14b) respectively would look like on M&N's account.

---

7.   For a discussion of further variation regarding Imp_{se} + *lo* sequences, not just in Spanish but in other Romance languages, see Mendikoetxea (1999); Mendikoetxea & Battye (1990); Ordóñez & Treviño (2016), and MacDonald & Melgares (2021).

8.   There is a repair for these sequences in Honduran Spanish which changes *lo* to *le*, but only in the matrix context: *Se le quiere abrazar.* **Se quiere abrazarle.* This is another asymmetry that we do not discuss here in detail.

(14) a. *Matrix Clause (13a)*
… [$_{TP}$ **Imp$_{se}$-lo**-T [$_{vP}$ <**Imp$_{se}$**> [$_{VP}$ quiere…<lo>]]]

b. *Embedded Clause (13b)*
… [$_{TP}$ **Imp$_{se}$** T-**lo** [$_{vP}$ <**Imp$_{se}$**> [$_{VP}$ abrazar…<lo>]]]]

The same conditions appear to hold in both the matrix and embedded clause, on the assumption that Imp$_{se}$ originates in the embedded clause. Only in the matrix context, however, is the sequence of Imp$_{se}$ plus *lo* ungrammatical. This asymmetry is not expected on M&N's account.

The ungrammatical Imp$_{se}$ plus *lo* sequence is clearly not an identity violation, but a morphosyntactic constraint on clitic sequences. Importantly, on M&N's account identity violations are sensitive to morphological information. Consequently, there is no obvious reason why the same symmetrical matrix-embedded clause effects should only be limited to sequences of identical clitics.

Like Honduran Spanish, European Portuguese exhibits the patterns in (13), as observed by M&N. They call it an adjacency effect, which suggests a different explanation for the patterns. We believe that this assumption is not warranted, at least for Spanish, since there is evidence from Spanish that the ungrammatical sequence of Imp$_{se}$ plus *lo* is not strictly a linear adjacency effect. This comes out when considering the sequence in ditransitive configurations. Consider the ditransitive in (15a) and its impersonalized form in (15b).

(15) a. *Me   recomendaron a   Juan.*
To.me recommended DOM Juan
"They recommend Juan to me."

b. *Se   me   recomienda a   Juan siempre.*
Imp$_{se}$ to.me recommend DOM Juan always
"They always recommend Juan to me."

Pronominalization of *Juan* in (15a) and (15b) gives us (16a) and (16b) respectively. (See Calandria, Palací & Saab 1999 for data parallel to 16b.)

(16) a. *Me   lo   recomendaron.*
To.me him recommended

b. %*Se   me   lo   recomienda siempre.*   [√Riopl/*Hond][9]
Imp$_{se}$ to.me him recommend always

---

9.   Note that the *lo* to *le* repair available in Imp$_{se}$ constructions (see footnote 8) is not available here, because the result leads to a PCC violation which rules out sequences of *me le* in ditransitives: **se me le recomienda*. Note also that in *leísta* varieties which have available a PCC violation repair, which changes *le* to *lo* is not available here, because it leads to (16b) in the text body, which is ungrammatical. A repair of one leads to a violation of the other. If one of the violations were morphological in nature and one syntactic, we would expect that one might feed or bleed the other. Since this is not the case, they appear to be violations within the same domain, violations which are filtered out.

We know that we have the same effect of Imp_{se} plus *lo* that we saw from (13) above, since both Rioplatense and Honduran Spanish pattern in the same way. Specifically, both (16b) and (13) are grammatical in Rioplatense, while both are ungrammatical in Honduran Spanish.

Observe that the sequence of *se me lo* is not independently ungrammatical, as illustrated in (17) with aspectual *se* (Asp_{se}), ethical dative *me*, and direct object *lo*.

(17)   *Se   me lo han   comido.*
        Asp_{se} me it have eaten
        "They ate it up on me."

The important observation is that in (16b) *se* and *lo* are not strictly linear adjacent, thus, it cannot be an adjacency effect. This is another matrix-embedded asymmetry, unexpected on M&N's account.

Based on these patterns, we offer an alternative explanation: Imp_{se} never merges in the embedded clause. If Imp_{se} were never in the embedded clause, the asymmetrical effects follow immediately, since we only expect to see them in the matrix clause. In the next section, we discuss instances where clitic climbing cannot take place for other clitics, but can for Imp_{se}, a set of circumstances consistent with the idea that Imp_{se} does not merge in the embedded clause.

3.   Clitic climbing

Recall that a key assumption on M&N's analysis is that in Imp_{se} constructions *se* merges in Spec,v of the embedded verb and moves to the matrix clause. In this section, we discuss patterns of clitic climbing and we see that Imp_{se} behaves differently from other climbing clitics. Specifically, Imp_{se} appears never to be blocked, when others are. If, as we suggested above, Imp_{se} were never in the embedded clause, then, we could straightforwardly explain why Imp_{se} is never blocked: it never climbs. As we suggest, this is unexpected on M&N's account.

To begin, a reviewer notes that one should not expect Imp_{se} to climb like other clitics, since M&N adopt a movement theory of control. More specifically, Imp_{se} undergoes A-movement to Spec,v, which is not ruled out since on a bare phrase structural account, clitics are both maximal and minimal, and therefore may move as a maximal projection. Cliticization to T occurs after A-movement. We submit here that this should be the null hypothesis, namely, that all clitics should be treated the same – they move as maximal projections until they cliticize.[10] It is not clear that this idea should set Imp_{se} apart from other clitics.

---

10. If not, then clitics would have to be able to excorporate (as in Roberts 1991, or Martins 2000), which we assume is not allowed (see also Matushansky 2006; Harizanov 2014).

On the other hand, whether or not A-movement is involved may play a role in whether clitic movement is blocked or not. Before addressing this question, we provide instances where Imp_{se} is not blocked like other climbing clitics are.

It has been widely observed that a *wh*-word and negation can block a clitic from moving from an embedded clause into its superordinate matrix clause (see Luján 1980; Martins 2000; Ordóñez 2012, *inter alia*). These are illustrated in (18) and (19) below, where in the presence of the *wh*-word *cómo* and negation, respectively, clitics are forced to remain within the embedded clause.

(18) a. *Sé   cómo decirlo (en español).*
          Know <u>how</u> say.it (in Spanish).
          "I know how to say it (in Spanish)."
      b. **Lo sé   cómo decir.*
          It   know <u>how</u>   say.
          "I know how to say it."

(19) a. *Deben no verla.*
          Should <u>no</u> see.it
          "They shouldn't see it."
      b. **La deben   no ver.*
          It   should <u>no</u> see
          "They shouldn't see it."

When the *wh*-word or negation is removed, the clitics can climb to the matrix clause, as in (20) and (21).

(20) a. *Sé   decirlo (en español).*
          Know say.it   (in Spanish)
          "I know how to say it (in Spanish)."
      b. *Lo sé   decir.*
          It   know say
          "I know how to say it."

(21) a. *<u>No</u> deben   verla.*
          <u>No</u> should see.it
          "They shouldn't see it."
      b. *<u>No</u> la deben   ver.*
          <u>No</u> it should see
          "They shouldn't see it."

Observe in (22), however, that the presence of the *wh*-word and the presence of negation does not prevent Imp_{se} from appearing grammatically.

(22) a. *No se   sabe <u>cómo</u> decirlo*
         No Imp$_{se}$ know <u>how</u>   say.it
         "One does not know how to say it."
     b. *Se   debe   <u>no</u> verla.*
         Imp$_{se}$ should <u>no</u> see.it
         "One should not see it."

Consider another instance, where clitic climbing in general is blocked, while the presence of Imp$_{se}$ is not affected. This is under ellipsis. Observe in (23) that the clitic can remain within the embedded clause (23a) or can climb to the matrix clause (23b).

(23) a. *María intentó robarla.*
         María tried   to.steal.it
     b. *María la intentó robar.*
         María it tried   to.steal
         "María tried to steal it."

Observe in (24), that ellipsis prevents the clitic from climbing, where angled brackets indicate ellipsis.

(24) **Susana la intentó <robar> también.*
     Susana it tried   to.steal also
     "Susana tried also."

Imp$_{se}$ again patterns differently, as illustrated in (25).

(25) a. *Aquí, se   intentó robar   una joya.*
         Here Imp$_{se}$ tried   to.steal a   jewel
         "Here they tried to steal a jewel."
     b. *Ahí, se   intentó <robar una joya> también.*
         There Imp$_{se}$ tried   to.steal a   jewel also
         "There they tried also."

Imp$_{se}$ patterns differently than other clitics. It appears never to be blocked (on the assumption that it moves). Let us now consider the relevance of A-movement to whether or not movement of a clitic is blocked.

Martins (2000) observes that in object control in European Portuguese an object clitic controller is not blocked by negation, like other clitics are. She claims that it is the clitic's need for accusative case that allows it to climb and avoid being blocked – because it A-moves, it is not blocked. We offer similar object clitic controller facts from Spanish in (26), but in ellipsis contexts. Under a movement theory of control, *me* would A-move from Spec,v of the embedded verb *robar* to the matrix clause.

(26) a. *María intentó obligarme a robar una joya.*
      María tried   make.me to steal a   jewel
      "María tried to make me steal a jewel."
   b. *María intentó obligarme <a robar una joya>.*
      María tried   make.me to steal a   jewel
      "Maria tried to make me."

Note the contrast with the direct object clitic corresponding to the complement position of *robar* in (24) above. On M&N's account, that both the subject control clitic Imp$_{se}$ and an object control clitic *me* pattern the same regarding movement out of an ellipsis site, is not unexpected, on the assumption that they are undergoing A-movement. Importantly, note that the patterns receive an alternative explanation: neither Imp$_{se}$ nor *me* in object control are in the embedded clause to begin with. Thus, if the embedded clause is elided, they are unaffected.

Moreover, if A-movement were the key to moving out of an ellipsis site, observe the patterns in (27) and (28), in which both *wh*-movement and A-movement of a passive subject are blocked under ellipsis.

(27) a. *María intentó ser detenida por la policía.*
      María tried   to.be detained by the police
      "María tried to be detained by the police."
   b. **Susana intentó < ser detenida por la policía> también.*
      Susana tried   to.be detained by the police   also
      "Susana tried also."

(28) a. *¿Qué intentó robar María?*
      What tried   to.steal María
      "What did María try to steal?"
   b. **¿Qué intentó <robar> Susana?*
      What tried   to.steal Susana
      "What did Susana try?"

Note furthermore that movement of object clitic controller *me* can also be blocked when moving to a higher clause. Thus, observe ungrammatical (29), which contrasts with grammatical (26b).

(29) **Susana me intentó < obligar a robar> también.*
      Susana me tried   to.make to steal   also
      "Susana tried also."

On the present account, the contrast between (29) and (26b) arises because movement only takes place in (29), not in (26b).[11] The same rationale applies to the instances where the apparent movement of Imp$_{se}$ is not blocked: Imp$_{se}$ is merged in the matrix clause, thus, it is outside the ellipsis site and is not affected by the presence of a lower *wh*-word or negation.

4.   An alternative proposal and an extension

An alternative to the patterns from (1) is offered by MacDonald & Vázquez-Lozares (2021). They claim that Imp$_{se}$ lacks number, thus, when it co-occurs with paradigmatic reflexives, the paradigmatic reflexives remain unvalued for number leading to a crashed derivation. Observe in (30) that Inher$_{se}$ is paradigmatic.

(30) a. *Yo me   casé.*
         I   Inher$_{me}$ got.married
     b. *Tú   te   casaste.*
         You Inher$_{te}$ got.married
     c. *Ella se   casó.*
         She Inher$_{se}$ got.married
     d. *Nosotros nos   casamos.*
         We   Inher$_{nos}$ got.married
     e. *Vosotros os   casasteis.*
         You.PL Inher$_{os}$ got.married
     f. *Ellos se   casaron.*
         They Inher$_{se}$ got.married
         "I/you/she/we/you(pl)/they got married."

For reasons of space, we do not review arguments for why Imp$_{se}$ lacks number, but see MacDonald & Vázquez-Lozares (2021) for a detailed discussion. See also Bruhn de Garavito, Heap & Lamarche (2002) for claims that Imp$_{se}$ is a severely underspecified clitic. If Imp$_{se}$ in Spanish lacks number, then, even non-identical reflexives that appear with it should be ungrammatical, just as we saw in (2) above. Two further examples are provided below in (31).

---

11. Martins (2000) claims that in clitic climbing contexts, the feature that drives movement is related to specificity. Clitic movement is blocked in cases where the clitic has already checked its feature in a lower clause, thus, there is no other feature driving movement. On that account, checking specificity in a lower clause is associated with more clausal structure. It is not clear that that approach extends to the contrast between (26b) and (29).

(31) a. **Se   tiene vergüenza de sí   mismo.*
        Imp$_{se}$ has shame   of self same
        Intended: "One has shame of himself."
   b. **Aquí se   quiere hablar de sí   mismo siempre.*
        Here Imp$_{se}$ wants speak of self same   always
        Intended: "Here, one always wants to speak about himself."

5. A brief recap

We have argued that M&N's account of ungrammatical sequences of Imp$_{se}$ plus other *se* clitics, which relies on their identity, cannot be extended to Spanish, based on what we called matrix-embedded asymmetries with Imp$_{se}$ plus Spur$_{se}$ and Imp$_{se}$ plus direct object clitic *lo*. We have also discussed how Imp$_{se}$ is not blocked like other clitics in clitic climbing contexts. We claimed that Imp$_{se}$ simply does not merge in Spec,v of the embedded verb, contra M&N, which immediately explains these matrix-embedded asymmetries. Furthermore, we noted that an alternative proposal exits, in which Imp$_{se}$ is not specified for number. The consequence is that Imp$_{se}$ cannot value any other paradigmatic *se*'s number, which leads to ungrammatical sequences of Imp$_{se}$ plus other *se* clitics. In addition, the alternative proposal can straightforwardly be extended to the ungrammatical appearance of non-identical reflexive *sí* in Imp$_{se}$ constructions in Spanish.

Acknowledgements

References

Bonet, Eulalia. 1991. *Morphology after syntax: Pronominal clitics in Romance languages.* Cambridge, MA: MIT dissertation.

Bruhn de Garavito, Joyce, David Heap & Jacques Lamarche. 2002. French and Spanish se: underspecified, not reflexive. *Actes du Congrès de l'Association canadienne de linguistique* 2002. 42–54

Calandria, Cecilia, Esteban Palací & Andrés Saab. 1999. El problema de la organización argumental y la concordancia. Paper presented at Congreso Internacional la Gramática: Modelos, Enseñanza, Historia, organizado por el Instituto de Lingüística, Facultad de Filosofía y Letras (UBA), Buenos Aires, Argentina, August 4–7.

Cinque, Guglielmo. 1988. On *si* constructions and the theory of *arb*. *Linguistic Inquiry* 19(4). 521–581.

Cuervo, María C. 2013. Spanish clitic clusters: Three of a perfect pair. *Borealis: An International Journal of Hispanic Linguistics* 2(2). 191–220. https://doi.org/10.7557/1.2.2.2692

Dobrovie-Sorin, Carmen. 1998. Impersonal *se* Constructions in Romance and the Passivization of unergatives. *Linguistic Inquiry* 29(3). 399–437. https://doi.org/10.1162/002438998553806

Harizanov, Boris. 2014. Clitic doubling at the syntax-morphophonology interface: A-movement and morphological merger in Bulgarian. *Natural Language and Linguistic Theory* 32(4). 1033–1088. https://doi.org/10.1007/s11049-014-9249-5

Luján, Marta. 1980. Clitic promotion and mood in Spanish verbal complements. *Linguistics* 18. 381–484. https://doi.org/10.1515/ling.1980.18.5-6.381

MacDonald, Jonathan E. 2017. An implicit projected argument in Spanish impersonal and passive-*se* constructions. *Syntax* 20(4).353–383. https://doi.org/10.1111/synt.12146

MacDonald, Jonathan E. & Matthew Maddox. 2018. Passive *se* in Romanian and Spanish: A subject cycle. *Journal of Linguistics* 54(2). 389–427. https://doi.org/10.1017/S002222671700038X

MacDonald, Jonathan E. & Jeriel Melgares. 2021. On (un)grammatical clitic sequences in impersonal *se* constructions. In Grant Armstrong & Jonathan E. MacDonald (eds.), *Unraveling the complexity of SE*, 87–110. Dordrecht: Springer. https://doi.org/10.1007/978-3-030-57004-0_4

MacDondald, Jonathan & Almike Vázquez-Lozares. 2021. Spanish impersonal *se* in control infinitivals and the ungrammaticality of *se se* sequences. In Tim Gupton & Elizabeth Gielau (eds.), *East and west of the Pentacrest: Linguistic studies in honor of Paula Kempchinsky*. John Benjamins. https://doi.org/10.1075/ihll.33.7mac

Martins, Ana María. 2000. A minimalist approach to clitic climbing. In João Costa (ed.), *Portuguese syntax: New comparative studies*, 169–190. Oxford: Oxford University Press.

Martins, Ana María & Jairo Nunes. 2017. Identity avoidance with reflexive clitics in European Portuguese and the theory of control. *Linguistic Inquiry* 48(4). 627–649. https://doi.org/10.1162/LING_a_00256

Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37(1). 69–109. https://doi.org/10.1162/002438906775321184

Mendikoetxea, Amaya. 1999. Construcciones con *se*: Medias, Pasivas e Impersonales. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática Descriptiva de la Lengua Española*, 1631–1722. Madrid: Espasa-Calpe.

Mendikoetxea, Amaya. 2008. Clitic impersonal constructions in Romance: Syntactic features and semantic interpretation. *Transactions of the Philological Society* 106(2). 290–336. https://doi.org/10.1111/j.1467-968X.2008.00210.x

Mendikoetxea, Amaya & Adrian Battye. 1990. Arb SE/SI in transitive contexts: A comparative study. *Rivista di Grammatica Generativa* 15. 161–195.

Nevins, Andrew. 2007. The representation of third person and its consequences for Person-Case effects. *Natural Language and Linguistic Theory* 25. 273–313. https://doi.org/10.1007/s11049-006-9017-2

Ordóñez, Francisco. 2012. Clitics in Spanish. In José Ignacio Hualde, Antxon Olarrea & Erin O'Rourke (eds.), *The handbook of Hispanic linguistics*, 423–451. Malden, MA: Wiley-Blackwell. https://doi.org/10.1002/9781118228098.ch21

Ordóñez, Francisco. 2021. On the nature of the impersonal SE: Why Italian is not like Catalan and Spanish. In Grant Armstrong & Jonathan E. MacDonald (eds.), *Unraveling the complexity of SE*, 137–160. Dordrecht: Springer. https://doi.org/10.1007/978-3-030-57004-0_6

Ordóñez, Francisco & Esthela Treviño. 2016. Agreement and DOM with impersonal se: A comparative study of Mexican and Peninsular Spanish. In Mary Kato & Francisco Ordóñez (eds.), *The morphosyntax of Portuguese and Spanish in Latin America*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190465889.003.0010

Pancheva, Roumyana & María Luisa Zubizarreta. 2017. The Person Case Constraint: The syntactic encoding of perspective. *Natural Language & Linguistic Theory* 36(1). 1291–1337. https://doi.org/10.1007/s11049-017-9395-7

Perlmutter, David. 1971. *Deep and surface structure constraints in syntax*. New York: Rinehart and Winston Inc.

Pescarini, Diego. 2011. Mapping Romance clitic sequences. *Quaderni di lavoro ASIt* 12. 1–29.

Roberts, Ian. 1991. Excorporation and minimality. *Linguistic Inquiry* 22, 209–218.

Walkow, Martin. 2012. Goals, big and small. Amherst, MA: University of Massachusetts Amherst dissertation.

On the interpretation of the Spanish 1st person plural pronoun

Ángel L. Jiménez-Fernández and Mercedes Tubino-Blanco

Universidad de Sevilla / Western Michigan University

We explore the connection between the clause left periphery and the referential values of inclusion or exclusion of the addressee associated with Spanish 1st Person Plural pronoun *nosotros*, otherwise morphologically marked in many world languages. We examine the referential values of *nosotros* in clauses marked with different topics and foci regarding the inclusive/exclusive interpretation of the pronoun. We observe, contra Posio (2012), that overt *nosotros* doesn't always involve exclusivity. The exclusive interpretation of the pronoun is nonetheless required in contexts typically declarative and non-contrastive (i.e., out-of-the-blue, thetic), and its overt use is perceived as odd if the Addressee is intended to be included. In Aboutness-Shift Topic and Given Topic contexts the clusive interpretation of the pronoun is obtained from the immediate context (i.e., whether the Addressee is active). Similarly, contrastive topics or foci include the presence of the Addressee in the immediate context as one of their points of contrast. To account for the influence of the immediate context on the interpretation of the pronominal values of clusivity we propose an analysis based on the projection of a Speech Act Phrase (SAP) (Speas & Tenny 2003) in combination with a Logophoric Center (Bianchi 2003) above the clausal left periphery (CP). In our analysis, the pronoun *nosotros* has an [Addressee] feature that is valued according to the availability of the Addressee in the left-most left periphery. Our formalization of clusivity assumes that interpretation is read off syntax (Haegeman & Hill 2013).

Keywords: clusivity, 1st person plural reference, left periphery, speech act syntax, logophoricity

1. Introduction

In this chapter, we explore the connection in Spanish between the clause left-most left periphery and the interpretation of the referential values of inclusion/exclusion of the Addressee associated with 1st person plural (1PL), as expressed by the overt

pronoun *nosotros*. While these referential values are explicitly marked in the morphology of many world languages (Cysouw 2002), they are not realized morphologically in Spanish. However, Spanish clusivity has been observed (Posio 2012) to have a potential impact on Information Structure (IS), as exclusive 1PL reference was claimed to be associated with overt *nosotros*, whereas inclusive reference was realized by a null pronoun.

In order to determine whether the 1SG pronoun is in fact associated with IS, we examine referential values of *nosotros* in sentences marked with different types of topic (Frascarelli & Hinterhölzl 2007), and foci (Jiménez-Fernández 2015), especially regarding inclusive/exclusive restrictions.

Contra Posio (2012), we argue that it is not the overt/null nature of the pronoun (i.e., IS) that determines its interpretation, but the answer is in the left-most left periphery, which determines the reference of indexicals (Bianchi 2003; Delfitto & Fiorin 2014).[1] More precisely, the inclusive or exclusive interpretation of 1PL is due to a mechanism that anchors the deictic components of an event to the context of Speech. While 1PL comes from the lexicon underspecified for its Person reference, its interpretation is anchored to the context by means of logophoric features housed in the Logophoric Center (Bianchi 2003), in the left-most left periphery. Part of the information provided by logophoric features is the presence of the Addressee in the interpretation of a pronoun, as available in the discursive context. The nature of the Speech Act (SA) (e.g., interrogative, imperative) may also naturally presuppose or exclude the Addressee. We adopt the Speech Act (SA) system proposed by Speas (2000) and Speas & Tenny (2003), formalized in terms of a Speech Act Phrase, immediately dominated by the Logophoric Center, and dominating the left periphery of the clause. If the logophoric operator in DiscourseP or the nature of SA values the [+Adressee] in SA, 1PL will also value this feature in CP, TP and vP, accounting for its inclusive interpretation. Briefly, the Logophoric Center is instantiated in our analysis by projecting a SAP above CP. This SAP is in turn dominated by the DiscourseP. Any relation between a pronominal and the clusivity feature in SAP is viewed as an Agree relation.

The use of a SAP projection to account for the interpretation of pronominal features has been independently adopted by D'Alessandro (2007) to describe the person feature of *si* in Italian, and by Woods (2014) to account for the anchoring of speaker oriented adverbs with the Speaker or Addressee.

---

1.   For Bianchi (2003) indexicals are the values assigned to the participants in a communicative act by means of a semantic anchoring with the Speaker or the Addressee activated in the Logophoric Center. In our system, these indexicals are viewed as a consequence of an AGREE relation with the Speaker and/or the Hearer as activated by the Logophoric Center.

The paper is organized as follows: in § 2 we present some background and data concerning the use of subject pronouns in Spanish and more specifically the 1PL reference. In § 3 we put forth the proposal that IS doesn't determine clusivity or logophoricity, but rather this interpretation depends on the context. In § 4 a formal analysis of the inclusive or exclusive interpretation of 1PL pronoun is proposed based on a SAP in the leftmost part of the left periphery. Section 5 presents our conclusions.

2.   Background and data

2.1   The use of Spanish subject pronouns and their IS

Overt subject pronouns in Romance languages such as Spanish, Portuguese and Italian are traditionally characterized as occupying an IS-based position in the Left Periphery (Camacho-Taboada, Jiménez-Fernández & López-Rueda 2016). Spanish, as a null subject language, doesn't require overt subjects if their deictic information is evident from discourse (Fernández-Soriano 1999). This is illustrated in (1) where the null subject is represented as Ø:

(1)   Ø *Trabajamos* de 9–5 todos los días.
      "We work from 9–5 every day."   Fernández-Soriano (1999: 1224 [48d])

However, subject pronouns are necessarily overt in very specific contexts, even if verbal morphology unambiguously shows the subject reference, for example, to indicate switch reference (de Cock 2011):

(2)   *En casa **mi marido** friega los platos porque **yo** odio hacer eso.*
      "At home my husband does the dishes because I hate doing that."
                                          Fernández-Soriano (1999: 1227 [56e])

Current line of research on pronouns (e.g., *NGRAE* 2009; Camacho-Taboada et al. 2016) associates overt subject pronouns with particular discourse uses as topics or foci. Camacho-Taboada et al. (2016), based on types of topics (Frascarelli & Hinterhölzl 2007) and foci (Jiménez-Fernández 2015), analyze the specific IS function of subject pronouns with respect to their overt or null realization:[2]

---

2.   We give a full description of types of topics and foci in § 3.

a. *Topics*

Aboutness-shift (AS-) topics are optionally overt, as shown in (3), Contrastive (C-) topics are obligatorily overt, as shown in (4), and Given (G-) topics are typically null, as seen in (5) – Examples (3–8) are from Camacho-Taboada et al. (2016):[3]

(3)   *Has estado hablando de Juan durante horas.… Me han dicho que (él) no sabe nada de los resultados del examen.*[4]
"You've been talking about Juan for hours… I heard that (he) doesn't know anything about the exam results."

(4)   A: *¿Cómo nos organizamos para preparar la fiesta?*
"What are each of us doing to organize the party?"
B: **(Yo) me encargo de la compra, *(tú) puedes enviar las invitaciones.*
"I'll do the shopping, you can send out the invites."

(5)   *No he visto a María desde mayo. (*Ella)/pro debe estar muy ocupada.*
"I haven't seen Maria since May. She must be very busy."

b. *Foci*

While Mirative (M-) foci may be both null and overt, as seen in (6), Contrastive (C-) and Information (I-) foci are obligatorily overt, as shown in (7) and (8) respectively:[5]

(6)   *¡No puedo creerme eso de María! ¡(ELLA) ha terminado sus estudios de doctorado!*
"I can't believe that about Maria! SHE finished her PhD!"

(7)   A: *He organizado todo para la fiesta de cumpleaños de Jimena.*
"I have organized everything for Jimena's birthday party"
B: *No, no. *(Yo) he organizado todo, no *(tú).*
"No, no. I have organized everything, not you."

(8)   A: *¿Quién ha roto el vaso?*
"Who broke the glass?"
B: *Ha sido *(ELLA).*
"It was HER."

---

3.   See Pešková (2014) for examples of Given Topics with overt pronouns.

4.   The pronoun in this example is not contrastive, but refers to a participant (i.e., *Juan*) previously mentioned.

5.   The optionality of the pronoun in the case of M-Focus interpretation is crucially influenced by the fact that either only the pronoun develops the M-Focus function or mirativity affects the whole sentence. In the former reading the pronoun is obligatory whereas in the latter it is not.

2.2 The pronominal interpretation of discourse participants

Concerning the deixis of pronouns in discourse, Benveniste (1966) states that 1st and 2nd person reflect a reality of discourse (i.e., Speaker and Addressee), and they are reversible in the communicative act, since speakers and addressees change their reference. For Benveniste, 1PL is not the plural of 1SG, but rather the combination of 1SG (the speaker) and others, its reference being unspecified.[6]

1PL is complex in its deictic reference (Harley & Ritter 2002), and naturally associated with several interpretations, hence its referential vagueness, as discussed in Posio (2012), and Di Tullio (2016). Noyer (1992) and Harley & Ritter (2002) define the reference of 1PL in terms of whether the Addressee is presupposed in the feature composition of the pronoun. Harley & Ritter (2002) formalized the reference of 1PL in terms of the hierarchy of features shown in (9). As seen in the diagrams, the main difference between inclusive and exclusive reference has to do with the presence/absence of the Addressee as a Participant encoded as part of the feature composition of the pronoun: Its presence is associated with inclusive reference whereas the Addressee is absent as a discourse participant in the case of exclusive reference:

(9) Feature specification of pronouns (Harley & Ritter 2002)
    a. Inclusive

```
                RE (Referential Expression)
                  /              \
          Participant          Individuation
           /      \                 |
     Speaker    Addressee        Group
```

    b. Exclusive

```
              RE (Referential Expression)
               /              \
        Participant         Individuation
            |                    |
         Speaker              Group
```
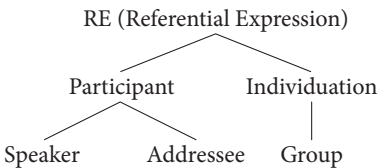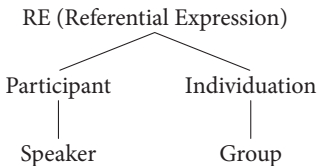
2.3 The expression of clusivity

Many world languages (35% of the languages considered by Harley & Ritter 2002) including many Australian and Amerindian languages (Cysouw 2002), mark the inclusive/exclusive distinction morphologically. This is illustrated in (10) for Guarani, as described in Di Tullio (2016):

---

6. See also Harley & Ritter (2002) and Di Tullio (2016).

(10)   Guarani
a.   *yané* (1+2(+3))
b.   *oré* (1+3)   Di Tullio (2016)

In Spanish clusivity is morphologically unspecified, but 1PL is still compatible with multiple readings, containing reference to the speaker plus a hearer or a non-participant, as we can see from the grammaticality of all sentences in (11).

(11)   a.   *Mañana no todos vamos a esquiar; tú no vas.*
            "We're not all going skiing tomorrow; you're not going."
       b.   *Mañana no todos vamos a esquiar; yo no voy.*
            "We're not all going skiing tomorrow, I am not going."
       c.   *Mañana no todos vamos a esquiar; Juan no va.*
            "We're not all going skiing tomorrow, Juan is not going."

The sentence in (11a) negates the interpretation of the addressee in the reference of 1PL (e.g., *tú*), (11b) negates the interpretation of the speaker (e.g., *yo*), and (11c) negates the interpretation of a non-participant (e.g., *Juan*).

    Although clusivity is morphologically unmarked in Spanish, inclusive interpretations may be required by constructions that involve an Addressee by default (e.g., imperative) or conditioned by context-anchoring adverbials (e.g., *entre tú y yo* "between you and me"):

(12)   a.   ***Vayamos** a la fiesta del sábado.*
            "Let's go to the party on Saturday."
       b.   ***Entre tú y yo** tendremos muchos invitados.*
            "Between you and me we'll have many guests."

Conversely, exclusive interpretations may be obtained by different means. For example, the Speaker in a conversational opening move may exclude herself from being an event Participant in the reply by using 2PL reference. This would trigger an exclusive interpretation if 1PL is used:

(13)   A: *¿Qué **hacéis**$_{[+A/-S]}$ este sábado?*[7]
            "What are you all doing this Saturday?"
       B: *Vamos$_{[+S/-A]}$ a una fiesta. ¿Te <u>apuntas</u>$_{[+A]}$?*   (exclusive)
            "We're going to a party. Wanna come?"

The referential interpretation of 1PL in terms of clusivity has not been explored for Spanish in detail. Based on his corpus, Posio (2012) associates the overt use of

---

7.   Indices referring to the presence [+] or absence [−] of an Addressee [A] or a Speaker [S] are used throughout the chapter for ease of exposition.

nosotros with exclusive reference (14). According to this author, only null pronouns may be interpreted inclusively (15):

(14) ***Nosotros**$_{[-A]}$ nos encargamos del trasplante de medula ósea pero previamente otros compañeros pediatras se han ocupado de la niña en los primeros momentos.*
"**We** are in charge of the bone marrow transplant but other colleagues have previously treated the child." Posio (2012: 349[10])

(15) *… si le parece doctora, ø nos va**mos**$_{[+A]}$ acercando a la burbuja donde está Elena, mientras tanto.*
"If you'd like Dr., we can start walking to the sterile zone where Elena is, in the meantime." Posio (2012: 350[12])

However, further data from Spanish shows that the overt use of *nosotros* is not limited to exclusive interpretations, as seen in (16) where the interpretation of the overt pronoun may be inclusive:

(16) A: *Juan y Antonio van a Japón este año.*
"Juan and Antonio are going to Japan this year."
B: *¿Y **nosotros**$_{[+A]}$ no?*
"And we're not?"

Moreover, the interpretation of an overt 1PL pronoun necessarily feeds from contextual information (including gestures, see Ortega-Santos 2016), as its reference in (17) is compatible with both readings:

(17) A: *Juan y Antonio van a Japón.*
"Juan and Antonio are going to Japan."
B: *¿Y **nosotros**$_{[+/-A]}$ no?* (inclusive or exclusive)
"And we're not?"

From examples (16–17) we conclude that overt pronouns cannot be limited to an exclusive interpretation, as claimed in Posio (2012). In what follows, we show that the immediate discourse context needs to be considered to draw 1PL pronoun clusivity.

3. Information structure does not determine clusivity

Contra Posio (2012), we claim that 1PL is referentially unspecified, either overt or null, and depends on the left-most left periphery to determine its Person reference. More precisely, the logophoric interpretation of 1PL depends on (i) the immediate context and the Participants' shared knowledge (i.e., Common Ground), and (ii) the role of discourse Participants as established by the Speech Act.

Information Structure also feeds from this information, but it does not directly determine the reference, which explains why the overt nature of person features isn't typically associated with a particular interpretation.

3.1 Topics

In § 2 above we saw that whereas AS-Topics are optional, C-Topics are obligatory and G-topics are typically null. In this section we show that 1PL topics may be associated with either kind of interpretation regardless of its subclass.

3.1.1 *Aboutness-Shift Topics*

AS-Topics newly propose or reintroduce a topic. In this IS function, the person reference of 1PL must match the Participants already established by the previous context. Our prediction is that 1PL AS-Topics may be interpreted inclusively or exclusively, depending on the Person features available from the immediate context. This is borne out in (18) and (19), with the inclusive interpretation of *nosotros* retrieved from the contextual Speaker plus the Addressee (18), and the exclusive interpretation in (19) made available by the immediate reference of a Speaker plus a Non-Participant (i.e., *Juan*):

(18) *Has estado hablando <u>de ti y de mí</u> durante horas…. La verdad es que (**nosotros**$_{[+A]}$) nunca nos había**mos** llevado tan bien.*
"You've been talking about you and me for hours… You can really tell we had never got along so well."

(19) *Siempre me preguntas <u>por mí y por Juan</u>. **Nosotros**$_{[-A]}$ nunca hemos estado unidos.*
"You're always asking about me and Juan, but we have never been close."

3.1.2 *Contrastive Topics*

According to Lee (2006), C-Topics are given, presupposed or anchored in the speech situation, just like non-contrastive Topics, but they are marked since they break down members of a superset previously established.

As with AS-Topics, 1PL C-Topics may be interpreted as inclusive (20), combining Speaker and Addressee, and also exclusive as in (21):

(20) A: *¿Cómo nos organiza**mos**$_{[+A]}$ para preparar el viaje?*
"What shall we each do to organize the trip?
B: **(**Nosotros**$_{[+A]}$) nos encargamos de los vuelos, <u>**ellos**</u> pueden buscar hotel.*
"We will look at the flights; they can search for a hotel."

(21) A: *¿Cómo nos organiza**mos**$_{[+A]}$ para preparar el viaje?*
"What shall we each do to organize the trip?
B: **(**Nosotros**$_{[-A]}$) nos encargamos de los vuelos, <u>**tú**</u> puedes buscar hotel.*
"We will look at the flights; you can search for a hotel"

3.1.3   *Given Topics*

G-Topics are used to (i) provide continuity with respect to the AS-Topic, or (ii) reprise background information. When pronominal, they are usually null in Spanish (Jiménez-Fernández 2016). G-Topics are compatible with both an inclusive (22) and an exclusive interpretation (23):

(22)   A: *¿Nos apuntamos*[+A] *tú y yo a la clase de yoga?*
             "Shall both you and I sign up for the yoga class?"
       B: *Vale, y (**nosotros**[+A]) tenemos descuento de estudiante.*
             "OK, and we have a student discount"

(23)   A: *¿Os apuntáis*[+A] *a tomaros unas cervezas?*
             "Are you down for some beers?"
       B: *¡Sí, sí! (A **nosotros**[−A]) nos encanta la cerveza!*
             "Yes! We sure love beer!"

3.2   Foci

Recall from § 2 that M-Foci are typically overt if mirativity concerns the pronoun alone and null if mirativity extends to the whole proposition. C- and I-foci are obligatorily overt. In this section we show that 1PL foci may be associated with either kind of interpretation regardless of its class, just like topics.

3.2.1   *Mirative foci*

For Cruschina (2011) and Jiménez-Fernández (2015), mirativity is new information that is unexpected, involving some sort of surprise for the speaker. It establishes a contrast with an element that is part of the Participants' shared knowledge (Bianchi, Bocci & Cruschina 2015). The logophoric interpretation of mirative focus 1PL pronouns can be either inclusive or exclusive, as the ambiguity of (24) suggests:

(24)   *¡No me lo puedo creer! ¡(**nosotros**[+/-A]) hemos entrado en los estudios de doctorado!*
        "I can't believe it! We've been accepted in the PhD program!"

3.2.2   *Contrastive foci*

For Zubizarreta (1998), C-foci denote a constituent that is clearly contrasted with another entity previously mentioned. The interpretation of deictics depends on the immediate context, including clusivity in 1PL reference. In (25) we find an example in which 1PL is interpreted as exclusive, since a contrast is established with reference to the Addressee. Inclusive reference is also possible if the contrast is made with a non-Participant (26):

(25) [A group of friends are talking about their respective vacation]
    A: *Bueno, <u>nosotros</u>[−A] hemos pasado unas vacaciones muy buenas.*
       "Well, we had a great vacation".
    B: *Pues* (**nosotros**[−A]) regular – María acabó en el hospital con gastroenteritis. ¡Que te cuente!*
       "We could have been better – Maria ended up in hospital with gastroenteritis. Ask her!"

(26) A: *Nadie lo ha pasado bien en la fiesta.*
       "Nobody had fun at the party".
    B: *Bueno,*(**nosotros**[+A]) no, pero <u>Juan</u> sí lo pasó bien.*
       "Well, WE didn't, but Juan did have fun."

3.2.3   *Information foci*

They denote purely new information (Zubizarreta 1998). Following Krifka (2006), in a question-answer exchange, I-Focus needs to satisfy the information search introduced by the *wh*-Phrase. If a 1PL pronoun is used, it needs to be overt and its reference will be determined by the information that is true for the Speaker. For this reason, it is unrestricted and it may be inclusive or exclusive:

(27) A: *¿Quién ha roto el jarrón?*
       "Who broke the vase?"
    B: *Lo hemos roto *(**nosotros**[+/−A]).*
       "WE did".

The logophoric interpretation of an I-Focus 1PL pronoun is strictly subject to contextual reference and difficult to predict from the linguistic context only.

3.3   Summary

The inclusive/exclusive reference of 1PL is not based on IS, since all information structure categories are compatible with both exclusive and inclusive reference. IS does seem to feed on Person specification available from the context, e.g., to continue reference in cases of some topics and to contrast reference to Person in cases of C-Foci.

Contra Posio (2012), the overt nature of the pronoun is not associated with clusivity or logophoricity, but it is rather meeting an exclusively discourse function as topic or focus. This leads us to claim that, in order to determine the inclusive/exclusive reference of the pronoun, we need to look at the Person specification available from the context as well as the nature of the Speech Act.

4. Speech act projection and the speaker/addressee relation

To interpret *nosotros* as inclusive, both discourse Participants need to be part of the specification of the pronominal reference, while exclusivity includes only the Speaker, as formalized in Harley & Ritter (2002). In the derivation we propose, the licensing of Person features for 1PL is multilayered, since the features that are interpreted in TP (i.e., Agreement) need to also be interpreted by IS in CP and very much depend on the nature of the Speech Act (e.g., imperative, interrogative, declarative), and the context of discourse (i.e., whether an Addressee is presupposed).

4.1 A logophoric center and a speech act phrase

To account for the anchoring of Discourse Participants to the syntactic derivation to determine the 1PL pronoun reference, we assume a superstructure that dominates the proposition: the Speech Act Phrase (Speas 2000; Speas & Tenny 2003). In our proposal, the Speech Act Phrase serves as a mechanism to anchor the deictic components of an event to the context of Speech (in DiscourseP). A 1PL pronoun comes from the lexicon underspecified (Di Tullio 2016). For interpretation, it needs to be anchored by logophoric features, as proposed by Bianchi (2003). In line with D'Alessandro (2007), [Speaker] and [Addressee] features are sheltered in SAP.
　　Bianchi (2003) proposes that every finite clause is anchored to a Logophoric Center, which is a speech or mental event (i.e., the utterance), with an obligatory animate Participant (the Speaker) and optional Addressee, as well as temporal and spatial coordinates. Following Bianchi (2003), 1/2 person features in pronouns are licensed via checking/valuing with SA in order to be deictically interpreted. But this value is determined by DiscourseP, the formal mechanism which in our system enables the interpretation of 1PL in terms of the Participants it encodes.[8] The Person feature is selected by [+finite] Fin°, in CP, which directly selects TP, where Person Agreement happens. The Logophoric Operator determines the contextual values of Person and its reference in CP, TP and vP. The diagram in (28) illustrates the derivation:

8.　As we mentioned at the outset, in our system Bianchi's Logophoric Center is interpreted as projecting a DiscourseP, whose specifier shelters a logophoric operator which may or may not activate the Addressee. This DiscourseP dominates SaP, which contains the relevant [Addressee] feature. This feature values the logophoric feature in pronouns as [+Addressee], thus obtaining the inclusive interpretation.

(28)
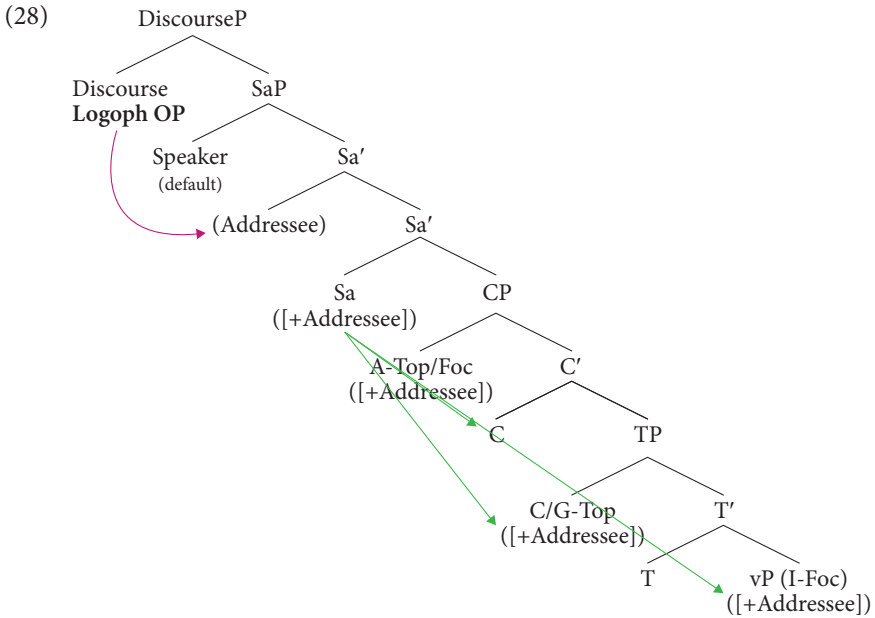
The logophoric operator determines whether the Addressee is present in the context of speech. The nature of the Speech Act (e.g., interrogative, declarative, imperative) may also require the presence of the Addressee in its Spec position. If the logophoric operator in DiscourseP or the nature of SA values the [+Addressee] in SA, 1PL will also value this feature in CP, TP and vP (depending on the type of discourse category), accounting for its inclusive interpretation. In other words, the logophoric operator acts as a probe in search of a goal, and finds the Addressee and possible occurrences of the unspecified 1PL, valuing their feature as [+Addressee], obtaining the inclusive reading.

4.2   Further evidence for SAP

Miyagawa (2017) offers evidence for the existence of an SAP from Jingpo, a Tibeto-Burman language that exhibits allocutive agreement. In Jingpo, agreement occurs on a sentence final particle, which may establish neutral agreement with the subject (29a), but agreement may also target the speaker, implying 'bonding', which results in 1PL agreement (29b).

(29) a. *Jongma du   hkum   m-s-ai*
          student arrive complete PL-PERF-3:DECL
      b. *Jongma du   hkum   sa-ga-ai*
          student arrive complete PERF-1PL-DECL
          "The students have all arrived"   (Miyagawa 2017)

Agreement with the speaker in combination with the clausal subject is also exhibited in the so-called empathic 1PL in Spanish (30), which combines the reference of the event participant in [Spec,vP] (i.e., the Addressee which has become active by the directive Speech Act, an interrogative) with the reference of the Speaker, a discourse participant, in [Spec,SaP]:

(30) *¿Cómo estamos$_{S+A}$ hoy?* (A mother to her kid)
      "How are we today?"

This strategy is generally involved in cases where the speaker includes self as a target for agreement along with another event participant, as part of a mitigation/solidarity strategy, as a way to express politeness, as in (30):[9]

(31) [gossiping with a friend about a co-worker wearing an extravagant dress]
      ¡Cómo venimos$_{-A+S}$ hoy!
      "How we're dressed today!"

Conversely, in cases of discursive strategies whereby the Speaker is augmented as group in the plural of modesty (Corbett 2000), the interpretation of 1PL is naturally exclusive, as the Speaker is the only event Participant, and an augmented Speaker gives way to exclusive interpretation in Harley & Ritter's (2002) analysis:

(32) *En este artículo planteamos$_{S+NP}$ …*
      "In this article we consider … "

In cases of unagreement as in (33), a DP unspecified for Person (e.g., *los lingüistas* "the linguists") may trigger 1PL agreement on the verb (e.g., *vamos* "go:1PL"). This is possible if we assume that verb agreement takes its reference from the left-most left periphery, where the discourse Participants (e.g., the Speaker) are encoded, as seen in the analysis in (28):[10]

(33) *Los lingüistas nos vamos$_{S+NP}$ de la sala.*
      "We linguists are leaving the room."

As a declarative in a thetic sentence, the sentence in (33) would typically be interpreted as exclusive. In imperative contexts, in contrast, 1PL is necessarily interpreted as inclusive (34), as imperatives involve [+Addressee] in SA:

---

9. We thank Maria Cristina Cuervo and Liliana Sánchez for examples along the lines of (31).
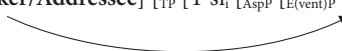
10. As seen in this chapter, Spanish does not mark clusivity morphologically. However, overt agreement with the Speaker as seen in (33) is morphosyntactic evidence in favor of positing a SAP in the left-most left periphery of the sentence, as the 1PL verbal agreement features cannot have been valued by the 3SG subject DP.

(34) *Vámonos$_{S+A}$ {nosotros$_{S+A}$ /los lingüistas$_{S+A}$} de la sala.*
"Let's/Let us linguists leave the room."

Politeness particles (e.g., *por favor* "please"), associated with Speech Act, force an inclusive reading, as they are interpreted as directives:

(35) a. *¿Vamos a la fiesta?*   (ambiguous)
"Are we going to the party?/Shall we go to the party?"
b. *¿Vamos a la fiesta, por favor?*   (inclusive)
"Shall we go to the party, please?"

Finally, D'Alessandro (2007: 170) argues that impersonal *si* constructions in Italian bounded by temporal adverbs receive an inclusive interpretation (i.e., their reference includes the Speaker). She proposes an analysis where the discourse participants are present in a Speech Act Phrase to value the Person feature on *si*:

(36) a. *Ieri si è arrivati tardi alla stazioni.*
"Yesterday we arrived late to the station."
b. [$_{SAP}$ [**Speaker/Addressee**] [$_{TP}$ [T si$_i$ [$_{AspP}$ [$_{E(vent)P}$ t$_i$]]]]]

All the phenomena just shown illustrates how, Discourse Participants in the left-most left periphery, combine their reference with event participants as targets for Agreement, justifying the need for both SAP and DiscourseP. In the next section we explain why overt 1PL pronouns are dispreferred in out-of-the-blue contexts if their interpretation is meant as inclusive.

4.3   Consequences in informatively unmarked contexts

Thetic or sentence-focus contexts are opening conversational moves (e.g., declarative out-of-the-blue statements), that introduce both new arguments and predicates (i.e., we can't have a topic):

(37) A: *Why didn't Mary come to work today?*
B: <u>*Her husband*</u> *is sick.*   Lambrecht (2000)

The prediction is that if we use a 1PL argument, it will be obligatorily interpreted as exclusive if an overt pronoun is used and the sentence is completely out-of-the-blue, since (i) overt pronouns are either contrastive or marked topics, (ii) 1PL establishes contrast in terms of clusivity, and (iii) an overt 1PL pronoun in a thetic clause searches the context for the Person reference it needs to complete its deictic content, and the Addressee is the only Participant available from the discourse context.

Hence in (38), a reply involving an overt 1PL and its default exclusive interpretation would be interpreted as odd:

(38)   A: *What's the plan?*
   B: [#]***Nosotros** vamos a la playa*. (if intended as inclusive)
         "We're going to the beach."

5.   Conclusions

It is the context that determines Person reference of 1PL in its relevant Speech Act. Information Structure does not determine Person reference, as 1PL may take either inclusive or exclusive value independently of its IS value and its null vs. overt nature.

IS also feeds from the leftmost periphery to draw Person reference, as this is a relevant point of contrast in the case of 1PL. In cases of contrast, 1PL plays on the two variables available in its inner composition (i.e., the presence or absence of the Addressee). If there is no contrast with a previously established value, the pronoun may be null if intended as topic.

These facts had never been formalized for Spanish as far as we know. This work sheds light on the 1PL reference and its relation with the left periphery. It also provides further evidence in favor of a Speech Act Phrase as necessary to capture the mechanisms by which pronouns obtain their Person reference from Discourse Participants. Overall, our analysis supports the view that discourse properties are present in the syntactic derivation (Haegeman & Hill 2013).

Gloss key

DECL	declarative
PERF	perfective
PL	plural

Acknowledgements

References

Benveniste, Émile. 1966. *Problèmes de linguistique générale*. Paris: Gallimard.

Bianchi, Valentina. 2003. On finiteness as logophoric anchoring. In Jaqueline Guéron & Liliane Tasmovski (eds.), *Tense and point of view*, 213–246. Paris: University of Paris X Nanterre.

Bianchi, Valentina, Giuliano Bocci & Silvio Cruschina. 2015. Focus fronting and its implicatures. In Enoch Aboh, Jeannette Schaeffer & Petra Sleeman (eds.), *Romance languages and linguistic theory 2013*, 3–20. Amsterdam: John Benjamins. https://doi.org/10.1075/rllt.8.01bia

Camacho-Taboada, Victoria, Ángel L. Jiménez-Fernández & Susana López-Rueda. 2016. The acquisition of null/explicit subject pronouns in Spanish as L2 by English speakers. In Bozena Cetnarowska, Marcin Kuczok & Marcin Zabawa (eds.), *Various dimensions of contrastive studies*, 310–332. Katowice: Wydawnictwo Uniwersytetu Slaskiego.

De Cock, Barbara. 2011. Why *we* can be *you*: The use of 1st person plural forms with hearer reference in English and Spanish. *Journal of Pragmatics* 43. 2762–2775. https://doi.org/10.1016/j.pragma.2011.04.009

Corbett, Greville. 2000. *Number*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139164344

Cruschina, Silvio. 2011. Fronting, dislocation and the syntactic role of discourse-related features. *Linguistic Variation* 11. 1–34. https://doi.org/10.1075/lv.11.1.01cru

Cysouw, Michael. 2002. Interpreting typological clusters, *Linguistic Typology* 6(1). 69–93. https://doi.org/10.1515/lity.2002.003

D'Alessandro, Roberta. 2007. *Impersonal "si" constructions: Agreement and interpretation*. Berlin: Mouton De Gruyter. https://doi.org/10.1515/9783110207514

Delfitto, Dennis & Gaetano Fiorin. 2014. Indexicality and left-periphery. *Rivista di Linguistica* 26(1). 59–92.

Di Tullio, Ángela. 2016. La primera persona del plural: una forma hipercaracterizada, talk delivered at *XIV Encuentro Internacional de Lingüística en el Noroeste*. Hermosillo.

Fernández-Soriano, Olga. 1999. El pronombre personal. Formas y distribuciones. Pronombres átonos y tónicos. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española* I, 1209–1274. Madrid: Espasa.

Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form*, 87–116. Amsterdam: John Benjamins. https://doi.org/10.1075/la.100.07fra

Haegeman, Liliane & Virginia Hill. 2013. The syntactization of discourse. In Raffaella Folli, Christina Sevdali & Robert Truswell (eds.), *Syntax and its limits*, 370–390. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199683239.003.0018

Harley, Heidi & Elizabeth Ritter. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78(3). 484–526. https://doi.org/10.1353/lan.2002.0158

Jiménez-Fernández, Ángel L. 2015. Towards a typology of focus: Subject position and microvariation at the discourse-syntax interface. *Ampersand* 2. 49–60. https://doi.org/10.1016/j.amper.2015.03.001

Jiménez-Fernández, Ángel L. 2016. When discourse met null subjects. *Borealis: An International Journal of Hispanic Linguistics* 5(2). 173–189. https://doi.org/10.7557/1.5.2.3727

Krifka, Manfred. 2006. Association with focus phrases. In Valeria Molnár & Susanne Winkler (eds.), *The architecture of focus*, 105–136. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110922011.105

Lambrecht, Knud. 2000. When subjects behave like objects: An analysis of the merging of S and O in Sentence-focus constructions across languages. *Studies in Language* 24(3). 611–682. https://doi.org/10.1075/sl.24.3.06lam

Lee, Chungmin. 2006. Contrastive (predicate) topic, intonation, and scalar meanings. In Chungmin Lee, Matt Gordon & Daniel Buring (eds.), *Topic and focus: Crosslinguistic perspectives on meaning and intonation*, 151–175. Dordrecht: Springer.

Miyagawa, Shigeru. 2017. *Agreement beyond phi*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/10958.001.0001

Noyer, Robert. 1992. Features, positions and affixes in autonomous morphological structure. Cambridge, MA: MIT dissertation.

Ortega-Santos, Iván. 2016. A formal analysis of lip-pointing in Latin-American Spanish. *Isogloss* 2(2). 113–128. https://doi.org/10.5565/rev/isogloss.41

Peškova, Andrea. 2014. Sujetos pronominales en el español porteño: Implicaciones pragmáticas en la interfaz sintáctico-fonológica [Pronominal subjects in Porteño Spanish: Pragmatic Implications at the syntax-phonology interface]. Hamburg, Germany: University of Hamburg dissertation.

Posio, Pekka. 2012. The functions of postverbal pronominal subjects in spoken Peninsular Spanish and European Portuguese. *Studies in Hispanic and Lusophone Linguistics* 5(1). 149–190. https://doi.org/10.1515/shll-2012-1122

Real Academia Española y Asociación de Academias de la Lengua Española. 2009. *Nueva gramática de la lengua española (NGRAE)*. Madrid: Espasa.

Speas, Peggy. 2000. Evidentiality, logophoricity and the syntactic representation of pragmatic features. *Lingua* 14. 255–276.

Speas, Peggy & Carol Tenny. 2003. Configurational properties of point of view roles. In Anna Maria Disciullo (ed.), *Asymmetry of grammar*, 315–344. Amsterdam: John Benjamins. https://doi.org/10.1075/la.57.15spe

Woods, Rebecca. 2014. The syntax of orientation shifting: Evidence from English high adverbs. In Martin Kohlberger, Kate Bllamy & Eleanor Dutton (eds.), *CONSOLE XXII: Proceedings of the 22nd Conference of the Student Organization of Linguistics in Europe*, 205–230. Leiden, Netherlands: The Leiden University Centre for Linguistics.

Zubizarreta, María Luisa. 1998. *Word order, prosody and focus*. Cambridge, MA: MIT Press.

Bridging issues at the PP-DP levels

French *ne … que* exceptives in prepositional contexts

J.-Marc Authier and Lisa A. Reed

The Pennsylvania State University

A common assumption in the literature is that French *(ne) … que* exceptives cannot appear inside a prepositional phrase. In this chapter we examine in some detail three contexts that counterexemplify this generalization. We show that in all cases, the data are consistent with the relatively recent hypothesis that *(ne) … que* exceptives contain a silent n-word (O'Neill 2011; Homer 2015; Authier & Reed 2019; Authier 2020). We also demonstrate that whenever an exceptive appears within a prepositional phrase, it is unable to take sentential scope. We provide three pieces of evidence that support this conclusion. First, PP-internal exceptives go hand in hand with the obligatory absence of the (optional) sentential negative scope marker *ne*. Second, PP-internal exceptives, unlike their PP-external counterparts, are unable to scope over modals in the clause in which they appear. Third, PP-internal exceptives, unlike their PP-external counterparts, do not license negative concord readings when they co-occur with a clause-mate n-word. These properties reveal the existence of a novel constraint that is not easily subsumed under existing theories of scope. That is, it seems that the combination of an exceptive's quantificational component (i.e., its n-word) with its association with focus property is such that any PP that contains it acts as a scope island.

Keywords: exceptives, focus, negative concord, prepositions, scope island

1. Introduction

Elaborating on ideas put forth in Baciu (1978), a number of researchers such as O'Neill (2011); Homer (2015); Authier & Reed (2019), and Authier (2020) have converged on the idea that French *(ne) … que* exceptives like (1) contain phonologically unrealized material. This material (crossed out in (1)) minimally consists of a silent n-word *rien* "nothing".[1]

---

1. O'Neill (2011) and Homer (2015) posit an additional silent component whose overt counterpart is *autre* "other". This second component is subsequently argued in Authier & Reed (2019) to

(1)   *Ils   (ne)   couperont ~~rien~~   que   les   érables.*
      they   (NEG) will-cut   nothing than the   maples
      "They will only cut down the maple trees."

The evidence in favor of assuming the presence of the silent n-word *rien* is quite robust. Besides the fact that it can appear overtly in many varieties of French, including colloquial standard French, Homer (2015) observes, for example, that when a *ne … que* exceptive and an n-word are clause-mates, they can give rise to either a negative concord (NC) reading (2a) or a double negation (DN) reading (2b). Since this behavior mimics what happens when two bona fide n-words are clause-mates, he concludes that *ne … que* has a covert n-word component.

(2)   a.   *Cette appellation ne   s'applique plus   qu'aux   vins*
           this   appellation NEG applies   no-longer than-to-the wines
           du   Languedoc.
           of-the Languedoc
           "This appellation now only applies to the wines of the Languedoc region."
           NC reading: This appellation no longer applies to any wine other than those wines.
      b.   *Ceci ne   s'applique plus   qu'aux   professeurs mais désormais*
           this NEG applies   no-longer than-to-the professors but   henceforth
           aussi aux   étudiants.
           also   to-the students
           "This no longer applies to just professors but from now on also to students."
           DN reading: This now applies to someone other than professors.

Thus, a reasonable conclusion seems to be that (3a) is the maximal phonological realization of a quantificational comparative which can also be spelled out as the minimal realization in (3b). That is, (3a–b) are syntactically and semantically equivalent and differ only in terms of those features that are accessed by the phonological component.

(3)   a.   *Je (n')ai   acheté rien   que des   tomates.*
           I   (NEG)have bought nothing than some tomatoes
      b.   *Je (n')ai   acheté ~~rien~~ que des   tomates.*
           I   (NEG)have bought   than some tomatoes
           "I only bought tomatoes."

---

be *plus* "more". However, Authier (2020) argues against positing a second silent element of this type on grounds that its overt counterpart does not display the association with focus property exhibited by overt or covert *rien* alone (see § 3 for examples).

This hypothesis immediately raises the question of whether the silent n-word represented as crossed out material in (3b) comes to be silent in the same way as a lexical element whose phonetic matrix is deleted at Spell-Out, instantiating PF-deletion ellipsis (see, e.g., Merchant 2001, 2004). To answer this question, let us take as a point of departure the feature-based taxonomy of lexical items given in Her & Tsai (2015). This taxonomy assumes that canonical lexical items have formal features (FF), which are accessible in the course of the narrow-syntactic derivation, as well as phonological features (PFF) and semantic features (LFF). While all lexical items active in syntax must have FF, non-canonical lexical items may lack PFF, LFF, or both. For example, overt expletives have no LFF, base-generated silent elements like PRO and *pro* have no PFF and null expletives have neither PFF nor LFF. Further, lexical items with no PFF, which we will simply call silent elements (SEs), differ from elements whose silence is due to ellipsis in that while the former have no PFF to begin with, the latter enter the derivation with PFF (i.e., as canonical lexical items) but their PFF are made invisible by ellipsis at the syntax-phonology interface (so-called PF deletion). A second important difference between SEs and elided elements has to do with the way in which they are subject to recoverability. That is, the meaning of SEs is recoverable from their pronounced counterparts in the lexicon, which means that SEs do not require overt discourse antecedents. PF-deleted elements, on the other hand, are recoverable only through overt discourse antecedents. Given this taxonomy, the silent component in (3b) must be seen as a SE rather than an elided element given that it is interpreted not via a discourse antecedent but, rather, by accessing the meaning of its PFF-endowed counterpart in the lexicon. Keeping these characteristics in mind, we devote the remainder of this chapter to the issue of availability of minimal exceptive *ne … que* under prepositions. As we show, while its syntactic derivation is consistent with the hypotheses formulated above, its semantic behavior as regards the availability of negative concord constitutes an unexpected puzzle which we carefully lay out and for which we offer some tentative explanations.

2.   Prepositional exceptives

Based on paradigms like (4), it has been widely assumed in the literature (see, e.g., Gross 1977: 90) that exceptive *que* can never follow a preposition.

(4) a. *Ce pays   ne compte **que sur**/***sur que** les revenus pétroliers.*
         this country NEG counts than on/on   than the revenues oil-related
         "This country relies only on oil revenues."

b. *On   ne   lit   ça   **que dans**/*__dans que__   cette feuille de chou.*
one NEG reads that than in/in   than this leaf   of cabbage
"You read this sort of thing only in this rag."

c. *Elle ne   voyage **qu'en**/*__en__   qu'avion.*
she NEG travels than-in/in than-plane
"She travels only by plane."

While the paradigm in (4) is representative of a wide range of examples, Damourette & Pichon (1943: 220) and Gaatone (1999: 105) point out that the generalization that has been drawn from it is not absolute in that it does not always hold in colloquial registers, as the examples in (5) show.

(5) a. *Faites trois séances si vous voulez, moi je ne   viendrai **à***
do   three sessions if you wish   me I NEG will-come to
***que* deux.**
than two
"Organize three sessions if you wish; me, I'll come to only two of them."
(M. ABA, July 2, 1919, recorded in Damourette & Pichon 1943: 220)

b. *[…] des   ministres **avec que** des   vraies factures […]*
some ministers with than some genuine bills
"ministers with only genuine bills"
(TF1, May 13,1991, recorded in Gaatone 1999: 105)

Furthermore, on closer inspection, the prepositional exceptive paradigm turns out to include more exceptions to the generalization based on examples like (4) than previously thought. Indeed, the data collected from our informants reveal that there are three patterns that deviate from the main paradigm in (4). First, there are instances, illustrated in (6), in which exceptive *que* not only may but must, in minimal realizations, follow the preposition regardless of register. These typically involve PPs headed by *avec* "with" that are complement to N. Further, such cases do not allow (optional) *ne*, suggesting that the scope of the exceptive is restricted to the PP.

(6) a. *Elle a   un verger **avec que** des   pommiers.*
she has an orchard with than some apple-trees

b. *__Elle a un verger **qu'avec** des pommiers.__*
"She owns an orchard with nothing but apple trees."

This scope restriction can be demonstrated by introducing a modal, which can only take wide scope with respect to the negation inherent to the covert n-word hypothesized to follow the preposition. Thus, (7a) must be understood as in (7b), not (7c). It is important to note that PPs headed by *avec* that are not complement to N, such as the one in (7d), behave differently. In such cases, exceptive *que* can appear before

the PP and be paired with *ne* and this results in the negation inherent to the covert n-word taking scope over the modal. Thus, (7d) receives the interpretation in (7e).

(7) a. *Elle peut acheter un verger **avec que** des   pommiers.*
          she can to-buy an orchard with than some apple-trees
     b. It is possible for her to buy an orchard with nothing other than apple trees.
     c. It is not possible for her to buy an orchard with anything other than apple trees.
     d. *On ne peut laver   ça   qu'avec   de l'eau.*
          one NEG can to-wash this than-with of the-water
     e. It is not possible for one to wash this with anything other than water.

Under the view adopted here that minimal realizations contain a silent n-word, the maximal realization of a sentence like (6a) is as in (8a) and the minimal realization is then obtained by using the SE counterpart to *rien*, yielding (8b). Thus, given the source in (8a), there is no way to derive the ungrammatical (6b), in which *que* precedes the preposition.

(8) a. *Elle a   un verger **avec rien   que** des   pommiers.*
          she has an orchard with nothing than some apple-trees
     b. *Elle a un verger **avec ~~rien~~ que** des pommiers.*
          "She owns an orchard with nothing but apple trees."

Next, there are maximal realizations, such as the one in (9), in which the preposition commanded by a verb may optionally appear before *rien* and must appear after *que*. The preposition involved in this pattern is usually *de* "of/about".

(9) *Ils (ne)   causent (**de**) rien   que *(**de**) politique.*
      they (NEG) talk   of   nothing than of   politics
      "They only talk about politics."

In such cases, the relevant paradigm for minimal realizations is as in (10).

(10) a. *Ils   ne   causent **de ~~rien~~   que de** politique.*
           they NEG talk   of nothing than of politics
      b. **Ils ne causent **de ~~rien~~ que de** politique.*
      c. **Ils ne causent **de ~~rien~~ que de** politique.*
           "They only talk about politics."

The grammaticality of (10a) follows from our assumption that exceptives contain a covert n-word, together with the fact that the first preposition in the source in (9) can be omitted. The ungrammatical examples in (10b–c), on the other hand, require that we add the further stipulation that the first preposition not only may

(as in (9)) but must be omitted when the n-word is covert. We will not attempt here to explain why this is so. We will simply note that providing an answer to this question requires that we first determine whether 'omitting' the first preposition means that it never enters the derivation or whether this preposition is syntactically present but phonologically silent.[2]

We turn next to the third and final paradigm, which consists of maximal realizations like (11) and (12). These feature a preposition that is commanded by a verb and can appear both before *rien* and after *que* (11a, 12a), just after *que* (11b, 12b), or just before *rien* (11c, 12c).

(11) a. ?*Je (n')ai   cuisiné **avec rien   qu'avec   des   produits   frais.***
          I (NEG-)have cooked with nothing than-with some ingredients fresh

      b. *Je (n')ai   cuisiné **rien   qu'avec   des   produits   frais.***
          I (NEG-)have cooked nothing than-with some ingredients fresh

      c. *J'ai   cuisiné **avec rien   que** des   produits   frais.*
          I-have cooked with nothing than some ingredients fresh
          "I cooked with only fresh ingredients."

(12) a. *Ce   sont (…) des   votes qui ne   débouchent **sur rien   que***
          these are   some votes that NEG lead   on nothing than
          sur la   désillusion.
          on the disillusion
          "These are votes that lead only to disillusionment."
                    (Excerpt from Valérie Précresse's speech during a political rally
          organized by François Fillon on April 9, 2017)

      b. *Ce sont (…) des votes qui ne débouchent **rien que sur** la désillusion.*

      c. *Ce sont (…) des votes qui débouchent **sur rien que** la désillusion.*

Given the paradigm in (11), we have three possible sources for minimal realizations. Assuming that the n-word in exceptives can be a SE, we then straightforwardly derive (13b) from the source in (11b) and (13c) from the source in (11c). Note that the fact that *ne* is optionally present in both (11b) and (13b) but obligatorily absent in both (11c) and (13c) gives further support to this analysis. Given the source in (11a), however, we incorrectly expect the minimal realization in (13a) to

---

2. Under pre-minimalist assumptions, one could have argued that silent categories do not need Case and that therefore a Case-assigning preposition is not needed when *rien* is covert. In the minimalist framework, however, Case valuation is seen as a by-product of ϕ-feature valuation under Agree, and the explanation is therefore bound to have far more reaching consequences for the theory at large. A further complication is that while a Google search on the sequence *de que de* "of than of" yields no hits, such is not the case for the sequence *avec qu'avec* "with than with" (see footnote 3).

be possible. Excluding it will again require that we stipulate that the first instance of the preposition that precedes *rien* must be omitted when the latter is covert.[3]

(13) a. **Je (n')ai   cuisiné **avec** ~~rien~~   **qu'avec** des   produits   frais.*
     I (NEG-)have cooked with nothing than-with some ingredients fresh

b. *Je (n')ai   cuisiné **rien**   **qu'avec** des   produits   frais.*
     I (NEG-)have cooked nothing than-with some ingredients fresh

c. *J'ai cuisiné **avec** ~~rien~~   que des   produits   frais.*
     I-have cooked with nothing than some ingredients fresh
     "I cooked with only fresh ingredients."

In sum, despite our stipulation concerning the ban on prepositions introducing covert *rien* in exceptives, our general assumptions concerning the nature of the full realization of minimal realizations allow us to generate all of the attested syntactic realizations of exceptive constructions in prepositional contexts. In what follows, however, we will show that things are not as straightforward at the syntax-semantics interface. Specifically, we will demonstrate that exceptives contained in a PP exhibit a behavior that diverge from their PP-external counterparts when it comes to the availability of negative concord readings.

3. Covert *rien* and negative concord in prepositional contexts

As is well-known, in French, when two n-words are clause-mates, they may (but need not) give rise to a negative concord (NC) reading whereby the semantic representation of the sentence contains one negation (cf. Déprez 1999, among many others). Thus, a sentence like (14a), which contains the two clause-mate n-words *jamais* "never" and *personne* "nobody", is ambiguous between a NC (single negation) reading (14b) and a DN (double negation) reading (14c).

(14) a. *Il n'y   a jamais personne dans ce   restaurant.*
     it NEG-there has never nobody in   this restaurant

b. There never is anybody in this restaurant.   (NC reading)

c. There always is somebody in this restaurant.   (DN reading)

---

3. Examples like (13a) are rejected by most speakers. They are, however, possible for some (in particular Belgian speakers), suggesting that there is some amount of dialectal variation. Examples like (i) are, in fact, relatively easy to find on the internet.

(i) %*Je fais moi-même mes savons   et   ne me lave   avec qu'avec   ça.*
     I make myself   my soap-bars and NEG me wash with than-with that
     "I make my own soap bars and I only use those when I wash."

A similar ambiguity arises in sentences with *jamais* in which the second n-word is contained in a PP, as in (15).

(15) a. *Adèle ne   danse   jamais avec personne.*
          Adèle NEG dances never   with nobody
      b. Adèle never dances with anybody.   (NC reading)
      c. Adèle always dances with somebody.   (DN reading)
          (i.e., Adèle never dances without a partner.)

Recall now that, as we saw in § 1, *ne … que* exceptives in non-prepositional contexts also give rise to these two readings when they appear in the same clause as an n-word. This, as we argued, follows from our hypothesis that the minimal realization of a *ne … que* exceptive contains a silent n-word. The relevant examples in (2) are repeated here as (16) for convenience.

(16) a. *Cette appellation ne   s'applique plus   qu'aux   vins*
          this   appellation NEG applies   no-longer than-to-the wines
          du   Languedoc.
          of-the Languedoc
          "This appellation now only applies to the wines of the Languedoc region."
          NC reading: This appellation no longer applies to any wine other than those wines.
      b. *Ceci ne   s'applique plus   qu'aux   professeurs mais désormais*
          this NEG applies   no-longer than-to-the professors but   henceforth
          aussi aux   étudiants.
          also to-the students
          "This no longer applies to just professors but from now on also to students."
          DN reading: This now applies to someone other than professors.

Given these facts, we then expect that sentences with an n-word like *jamais* "never" and an exceptive in a prepositional context should also be ambiguous between a NC and a DN reading, regardless of whether *rien* is overt or covert. This follows from our assumptions that (a) minimal realizations are syntactically indistinguishable from their maximal realization sources and (b) the SEs they contain are semantically recoverable by accessing the meaning of the PFF-endowed counterparts found in their maximal realizations. The facts, however, turn out to be more complex than expected. Although there is some dialectal (or perhaps idiolectal) variation, some clear patterns emerge from the judgements reported by a majority of our informants. The possible interpretations tied to maximal realizations are laid out in (17) and those exhibited by minimal realizations are given in (18).

(17) a. *Je (n')ai*   ***jamais*** *cuisiné **avec rien*** *qu'avec*   *des*
   I (NEG-)have never   cooked with nothing than-with some
   produits   *frais.*
   ingredients fresh
   AMBIGUOUS
   DN reading: "I've always cooked with something other than fresh ingredients."
   NC reading: "I've never cooked with anything other than fresh ingredients."

   b. *Je (n')ai*   ***jamais*** *cuisiné **avec rien***   *que des*
   I (NEG-)have never   cooked with nothing than some
   produits   *frais.*
   ingredients fresh
   UNAMBIGUOUS: *DN reading only*

   c. *Je (n')ai*   ***jamais*** *cuisiné **rien***   *qu'avec*   *des*
   I (NEG-)-have never   cooked nothing than-with some
   produits   *frais.*
   ingredients fresh
   AMBIGUOUS *(DN reading preferred)*

   d. *Je (n')ai*   ***jamais rien***   *cuisiné **qu'avec***   *des*
   I (NEG-)-have never   nothing cooked than-with some
   produits   *frais.*
   ingredients fresh
   AMBIGUOUS *(NC reading preferred)*

(18) a. *Je (n')ai*   ***jamais*** *cuisiné **avec que** des*   *produits*   *frais.*
   I (NEG-)have never   cooked with than some ingredients fresh
   UNAMBIGUOUS: *DN reading only*

   b. *Je (n')ai*   ***jamais*** *cuisiné **qu'avec*** *des*   *produits*   *frais.*
   I (NEG-)have never   cooked than-with some ingredients fresh
   AMBIGUOUS *(NC reading preferred)*

Given the interpretations displayed by the minimal realizations in (18), we arrive at a number of important conclusions. First, the maximal realization in (17a) cannot be the source for (18a), given the absence of a NC reading in the latter, neither can it be the source for (18b), which, although ambiguous like (17a), is reported by speakers to favor the NC reading, unlike (17a). Thus, deriving minimal realizations cannot be assumed to involve the use of silent (or PF-deleted) prepositions. Second, (17b) is clearly the source for (18a); that is (18a) differs minimally from (17b) in that it contains the silent version of the n-word in (17b). Finally, although (17c) and (17d) could in principle be the source for (18b), the preference for the NC reading exhibited by (18b) suggests that its most likely source is (17d), rather than (17c). Thus, (18b) differs minimally from (17d) in that it contains the silent version of the n-word *rien*. The proper representation for the paradigm in (18) is therefore as in (19).

(19) a. *Je (n')ai   jamais cuisiné avec ~~rien~~   que des*
         I (NEG-)have never   cooked with nothing than some
         produits   frais.
         ingredients fresh

     b. *Je (n')ai   jamais ~~rien~~   cuisiné qu'avec   des*
         I (NEG-)have never   nothing cooked than-with some
         produits   frais.
         ingredients fresh

These conclusions constitute, of course, further evidence in favor of the hypothesis that *ne … que* exceptives contain a silent n-word. However, they also highlight an unexpected difference between 'bare n-words' and n-words that are contained in exceptives. To explain, the generalization that arises from the paradigms in (17)–(19) is that the overt or covert n-word that is part of an exceptive located inside a PP cannot enter into a negative concord relation with another n-word higher in the clause that contains it. How do we then explain this constraint? One possibility is that, in exceptives, the n-word is part of a complex phrase and that this somehow blocks negative concord. The availability of a NC reading in examples like (20), however, appears to militate against this possibility.

(20) *Elle (ne) s'habille   jamais avec rien   de flamboyant.*
      she (NEG-) gets-dressed never   with nothing of flashy
      "She never wears anything flashy."   (*NC reading*)

A second possibility is that prepositional phrases constitute a boundary opaque to negative concord relations. While this is clearly true of some prepositions, for example *pour* "for" in (21), this cannot be true of *avec* "with" as the availability of a NC reading in (20) shows.

(21) *Elle (ne)   travaille jamais pour rien.*
      she (NEG) works   never   for   nothing
      DN reading only: "She always works for something."

On the other hand, it turns out that the lack of NC readings with exceptive n-words embedded in PPs is also found in quantificational comparatives in the same context. Consider in this respect the example in (22).

(22) *Je (ne) voyage jamais avec rien   de plus   que 20 euros en poche.*
      I (NEG) travel never   with nothing of more than 20 euros in pocket
      DN reading only: "I always travel with more than 20 euros in my pocket."

What this suggests is that French *ne…(rien) que* exceptives are bona fide quantificational comparatives, a conclusion argued for on independent grounds by Baciu

(1978); O'Neill (2011); Homer (2015), and Authier & Reed (2019), and for which Moignet (1973: 51) provides additional evidence of a diachronic nature. As Authier (2020) demonstrates, however, this characterization is insufficient because *ne…(rien) que* exceptives exhibit an association with focus property that is absent in their quantificational comparative counterparts. To see this, consider the paradigm in (23) (Authier's (57)).

(23) a. *Ce gâteau ne contient (rien) que du chocolat blanc.*
          this cake   NEG contains (nothing) than of-the chocolate white
          "This cake contains only white chocolate."
      b. *Ce gâteau ne contient rien de plus que du*
          this cake   NEG contains nothing of more than of-the
          chocolat blanc.
          chocolate white
          "This cake contains nothing more than white chocolate."

In (23a), both *ne … que* and *ne rien … que* associate with focus in a like manner. That is, while they take the constituent that follows them, namely, [*du chocolat blanc*] as their focus phrase, the focus they associate with can be the whole focus phrase or one of its sub-constituents, giving rise to ambiguity. If the focus coincides with the focus phrase, the resulting interpretation is that the proposition *this cake contains white chocolate* is true and any other proposition in the set of alternatives that includes propositions such as *this cake contains flour and eggs* is false; that is, one understands that the cake contains white chocolate and no other ingredient. If, however, the focus is a sub-constituent of the focus phrase, say, *blanc*, one understands that the proposition *this cake contains white chocolate* is true and any proposition in the set of alternatives made up of propositions like *this cake contains dark chocolate* is false; that is, the alternatives are varieties of chocolate rather than cake ingredients and the cake can therefore contain ingredients that are not chocolate as long as it contains white chocolate. This type of ambiguity displayed by *ne…(rien) que* is typical of items that associate with focus and is conspicuously absent from quantificational comparatives like *rien de plus que* in (23b). That is, the only interpretation associated with (23b) is that the cake contains white chocolate and nothing else. Thus, while it may very well be the case that French *ne…(rien) que* exceptives derived historically from quantificational comparatives, they have acquired along the way a focus sensitivity property that makes them semantically distinct. Whether this has implications for the syntax of exceptives ultimately depends on where in the grammar association with focus is assumed to be encoded. Focus in-situ appears to be insensitive to islands, as observed by Rooth (1996) in the context of focus particles like *only*. The same observation applies to *ne … que* sentences like (23a) where *blanc*, being a possible focus, occupies a

position inaccessible to syntactic movement. This suggests that the interpretation of exceptives does not involve covert syntactic movement, a conclusion consistent with Rooth's nonquantificational theory of focus. Focus can therefore be seen as a prosody-based, syntactically unencoded phenomenon. However, if association with focus has no syntactic correlates, it must be lexically encoded. While this is not a problem for lexical items like *only*, it is difficult to see how a lexical encoding account could apply to a multi-word expression like *ne … (rien) que*. This is in many ways reminiscent of the issue of how semantic presuppositions are encoded. That is, in most cases, the presence of presuppositions can be traced back to a single lexical item like a factive verb or a word like *again* but they sometimes arise in conjunction with specific syntactic configurations such as clefting. We will not discuss this issue any further here but will instead point out that it remains possible that *ne…(rien) que* exceptives and quantificational comparatives are syntactically identical and differ only semantically with respect to focus sensitivity.

4.   Concluding remarks

In this chapter, we uncovered three paradigms that contradict the generalization that French exceptives can never appear inside a PP. These paradigms were shown to support the claim that exceptives contain a silent n-word, which can also be overtly realized as *rien*. Further, in those cases in which an exceptive is contained in a PP, the n-word it contains cannot take sentential scope, as evidenced by the obligatory absence in such cases of the (optional) sentential scope marker *ne*. A direct consequence of this phenomenon is that exceptives embedded in a PP cannot, through their n-word component, participate in negative concord with a clause-mate n-word, unlike their non-embedded counterparts. Thus, exceptives in prepositional contexts reveal a new grammatical constraint not easily amenable to existing theories of scope. That is, the combination of a quantificational n-word component and the association with focus found in exceptives is such that any PP that contains them turns into a scope island. Why this should be so will be left as an open question for future research.

Acknowledgements

References

Authier, J.-Marc. 2020. On the comparative analysis of French *(ne) … que* exceptives. *Probus* 32. 1–54. https://doi.org/10.1515/probus-2019-0006

Authier, J.-Marc & Lisa Reed. 2019. Identifying phonologically overt counterparts to silent elements: The case of French exceptives. *University of Pennsylvania Working Papers in Linguistics* 25(1) 21–29.

Baciu, Ioan. 1978. La négation restrictive. *Le Français Moderne* 46. 135–142.

Damourette, Jacques & Edouard Pichon. 1943. *Des mots à la pensée. Essai de grammaire de la langue française [1911–1940] tome sixième*. Paris: d'Artrey.

Déprez, Viviane. 1999. The roots of negative concord in French and French based creoles. In Michel DeGraff (ed.), *Language creation and language change: Creolization, diachrony and development*, 375–428. Cambridge, MA: MIT Press.

Gaatone, David. 1999. Réflexions sur la syntaxe de *ne … que*. In Marc Plénat, Michel Aurnague, Anne Condamines, Jean-Pierre. Maurel, Christian Molinier & Claude Muller (eds.), *L'emprise du sens. Structures linguistiques et interprétations. Mélanges de syntaxe et de sémantique offerts à Andrée Borillo par un groupe d'amis, de collègues et de disciples*, 101–115. Amsterdam: Rodopi.

Gross, Maurice. 1977. *Grammaire transformationnelle du français. Syntaxe du nom*. Paris: Larousse.

Her, One-Soon & Hui-Chin Tsai. 2015. On silent elements: A case study of *grand* and its silent entourage. *Natural Language and Linguistic Theory* 33. 575–605. https://doi.org/10.1007/s11049-014-9266-4

Homer, Vincent. 2015. *Ne … que* and its challenges. In Ulrike Steindl, Thomas Borer, Huilin Fang, Alfredo García Pardo, Peter Guekguezian, Brian Hsu, Charlie O'Hara & Iris Chuoying Ouyang (eds.), *Proceedings of the 32nd West Coast Conference on Formal Linguistics (WCCFL 32)*, 111–120. Somerville, MA: Cascadilla Proceedings Project.

Merchant, Jason. 2001. *The syntax of silence: Sluicing, islands and the theory of ellipsis*. Oxford: Oxford University Press.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27. 661–738. https://doi.org/10.1007/s10988-005-7378-3

Moignet, Gérard. 1973. *Les signes de l'exception dans l'histoire du français*. Geneva: Droz.

O'Neill, Teresa. 2011. The syntax of *ne … que* exceptives in French. *University of Pennsylvania Working Papers in Linguistics* 17(1). 175–184.

Rooth, Mats. 1996. Focus. In Shalom Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, 271–297. London: Basil Blackwell.

Interpreting reduplicated numerals in Old Ibero-Romance

A syntactic account

Alice Corr
University of Birmingham

Old Ibero-Romance reduplicated numerals in transfer and possession constructions force a distributive reading at the sentential level (OSp. *los pecheros deben **tres tres** meajas* "the taxpayers owe **three** *meajas* **each**"). I argue that the distributive construction is best accounted for at the clausal level by an applicative structure (Pylkkänen 2008), where the reduplicated numeral is the internal argument and is encoded at the nominal level by a dedicated distributive layer in the extended functional structure. This analysis captures differences in the construction's interpretation – a low ApplP for distribution over individuals, and a high ApplP for distribution over events – within a more general theory of event structure, and situates the distributive construction with a wider taxonomy of (independently-motivated) Ibero-Romance applicative structures.

Keywords: distributivity, reduplicated numerals, applicatives, Old Ibero-Romance

1. Introduction

This chapter offers novel empirical data and a syntactic analysis of a construction found in Old Spanish (OSp), Old Portuguese (OPg) and West-Iberian Medieval Latin (IML) where a reduplicated cardinal numeral (*RedNum*) produces a distributive reading at the sentential level:

(1) … *mandóles   Moisén que tomassen   **dos dos medidas,***
    ordered.3SG=them Moises that take.IMPF.SBJV.3PL two two measures
    la una pora   esse día e   la otra pora'l sábado   (C13 OSp.)
    the one for.the that day and the other for.the sabbath
    "Moses ordered them to take **two omers [measures] each**, one for that day and the other for the Sabbath"
       (Horcajada Diezma & Sánchez-Prieto Borja, henceforth HD&SB, 1999: 281)

In (1), the distributive reading is obtained via the semantic relationship between a 'Distributive Share' (*DistShare*), an overtly-marked distributive expression (here, the RedNum ***dos dos medidas***, [*lit.* "**two two** omers"]), and a 'Distributive Key' (*DistKey*), the clausemate expression—here, the 3PL subject—over which the DistShare distributes its denotation (Gil 1988, 2005; Choe 1987; Zimmermann 2002). Originally observed in Ribeiro (1798), this construction, which I christen Old Ibero-Romance 'distance distributivity with numeral reduplication' ('DDNR'), is unattested elsewhere either historically or geographically in Latin or Romance. Whilst a modest body of studies has concentrated on a proposed Arabic/Mozarabic contact hypothesis (Sánchez-Prieto Borja & Horcajada Diezma 1994; Silva Villar 2015; see also Corr 2019: 84ff), no theoretical analysis of the phenomenon has yet been undertaken. This chapter thus not only builds substantially on existing work by offering newly-collected data of Old Ibero-Romance DDNR, but offers a unified formal analysis of the phenomenon that accounts both for the nature of the DDNR's surface morphosyntactic expression – viz. a requirement that distributivity be marked in the nominal structure, here via a non-deleted copy of the numeral – and for the sentence-level locus of its semantic effect.

Specifically, I propose and justify an applicative (Pylkkänen 2008) treatment of Old Ibero-Romance DDNR that captures (i) the defined set of clausal environments in which Old Ibero-Romance DDNR occurs, viz. transitive, ditransitive and unaccusative intransitive structures, where RedNum is always base-generated as the complement of V; (ii) its subject-object asymmetry; (iii) each construction's (individual or event) distributive reading; and iv) the lexical semantics of the relevant predicate. I propose that a double object-like construction, mirroring but distinct from the distributive functional structure hypothesised by Beghelli & Stowell (1997); Stowell (2013), and involving low or high Appl(icative) projections à la Pylkkänen (2008), underlies Old Ibero-Romance DDNR. This analysis accounts for differences in DDNR's interpretation – viz. a low ApplP when the DistShare distributes over individuals, and a high ApplP when it distributes over events – within a constructivist framework for argument/event structure, and unifies the construction with a wider taxonomy of (independently-motivated) applicative structures in Old and Modern Ibero-Romance.

1.1   Background

Typologically, reduplicated numerals are the most common strategy for encoding distributive numerals (*DistNum*), characterised as morphosyntactic constructions containing a numeral in which the sentence as a whole receives a distributive interpretation (Gil 2005: 222; Cable 2014: 563). Such items qualify as instances of DISTANCE DISTRIBUTIVITY (Zimmermann 2002) since the surface marking of

distributivity does not correspond to the locus of its semantic effect. In Romance and Germanic languages, distance distributivity is marked, optionally, via overt lexical strategies; namely, language-specific equivalents of 'shifted' (Postal 1974) or 'binominal' (Safir & Stowell 1988) *each*, a quantifier whose presence serves to enforce a distributive interpretation on an otherwise ambiguous sentence (2). Whilst 'binominal' *each* and its Romance equivalents can only distribute over individuals, grammaticalised DistNums (3) have a less restricted semantics (Balusu 2005):

(2)   *Las chicas compraron tres libros (**cada una**).*
      "The girls bought three books (**each [one.F]**)" (ModSp.)
    a.   Collective (unavailable with QP): one event of buying with multiple participants.
    b.   Distributive (obligatory with QP): multiple events of buying with one participant per event. (*For each occasion/location, there is an event of girls buying three books.)

(3)   ***Sypom-t.sypom-t*** *ombaky Ø-naka-'y-t*   *pikom.*   Karitiana
    two-ADJ.two-ADJ jaguar 3-DECL-eat-NFT monkey

                                  (Donazzan & Müller 2015: 122)
    a.   For each jaguar, there is an event of its eating two monkeys.
    b.   For each occasion/location, there is an event of jaguars eating two monkeys.

1.2   Data and sources

This study offers novel data for DDNR in Old Spanish, Old Portuguese and West-Iberian Medieval Latin retrieved via manual searches of *El Corpus Diacrónico del Español* (CORDE, corpus.rae.es/cordenet.html), *Corpus del Nuevo Diccionario Histórico del Español* (CNDHE; web.frl.es/CNDHE) and Barbosa Morujão (2010).[1] Secondary sources come from HD&SB (1999). I include West-Iberian Medieval Latin examples under the umbrella of 'Old Ibero-Romance' on the view that, despite its Latin morphology, its syntax is essentially that of Old Romance (cf. Wright 2010). The examples are attested in 10–15th century documents, mostly concentrated in the 13–14th centuries, and produced in central and northwest regions of the Iberian Peninsula, extending to La Rioja and Madrid in the East; and Toledo, Cáceres and Évora to the South.

---

1.   Results which form part of my corpus but are not illustrated in the present paper were obtained from *Corpus del español* (corpusdelespanol.org); *Corpus Informatizado do Português Medieval* (cipm.fcsh.unl.pt) and, from previous work, Ribeiro (1798) and Sánchez-Prieto Borja & Horcajada Diezma (1994). No results were obtained for Old Catalan from manual searches of the *Corpus Informatitzat del Català Antic*, corroborating the claim (HD&SB 1999: 295; Silva-Villar 2015: 192) that DDNR is unattested in this language.

2.   Syntactic environments

Old Ibero-Romance RedNums surface as an argument of the verb in the following clausal environments: ditransitive constructions (103 out of 125 tokens), where the RedNum/DistNum/DistShare is the direct object DP and the indirect object DP is the DistKey (4); transitive constructions (19 out of 125 tokens), where the RedNum is the object and the DistKey the subject (5); unaccusative intransitive constructions (3 out of 125 tokens), where, following Burzio (1986), the RedNum is assumed to be the internal argument (6); and, in one case, as a prepositional argument (7):[2]

(4) *mando   eisdem   archidiacono   et   magistroscolarum **tres   tres***
    send.1SG same.DAT archdeacon.DAT and schoolteacher.DAT three three
    marchas   de argento   (IML)
    half.pounds of silver
    "I leave to the above-mentioned archdeacon and schoolteacher **three half-pounds of silver each**"   (1291; Barbosa Morujão 2010: 188)

(5) *que trayan   cada cavero **tres   tres   bestias***   (OSp.)
    that bring.SBJV.3PL each worker three three beasts
    "Each worker should bring three beasts each"   (1251; CORDE)

(6) *entraron   a Noe al   arca **dos dos de toda criatura***   (OSp.)
    entered.3PL to Noe to.the ark two two of every creature
    "**Animals of every kind** went into the ark for Noah **two at a time**"
                                                         (c.1400; CORDE)

(7) *aparto-os   **com vinte   viinte** [sic] **soldos***   (OPg.)
    distribute.1SG=them with twenty twenty   soldos
    "I leave them **twenty *soldos* each**"   (1295; Barbosa Morujão 2010: 351)

Old Ibero-Romance DDNR displays subject-object asymmetry insofar as RedNum always originates in the same argumental position/thematic role, viz. the complement of V/THEME. In previous work (Corr 2018, 2019), I have shown that Old Ibero-Romance DDNR differs from the syntax-semantics both of RedNums/DistNums cross-linguistically (e.g., Balusu 2005; Cable 2014) and of binominal *each* (cf. §1.1). For reasons of space, I do not review the data here, but refer the interested reader to the above works.

---

2. Additionally attested in the historical record are a handful of DP-modifying RedNum PPs (OSp. *quarto sáuanas de **tres tres pannos*** "four sheets of **three three** cloths, i.e., each"), whose analysis I leave to future investigation.

3.   Accounting for Old Ibero-Romance DDNR

In this section, I present my proposal that the various syntactic constructions and lexical semantics involved in Old Ibero-Romance DDNR where RedNum is a verbal argument can be captured by a single (configurational) analysis. Departing from the observation that (i) DDNR is primarily attested in ditransitive structures (82.4% of tokens in the corpus) and (ii) all clausal constructions in which DDNR obtains involve a literal or figurative (transfer of) possession, I argue that Old Ibero-Romance DDNR constructions have a double object(-like) configuration. Whilst the literature offers two principal competing accounts for double object constructions (DOCs) – viz. a small clause analysis and an applicative analysis (cf. §3.3) –, I defend the position that the phenomenon is best accounted for in terms of an applicative structure à la Pylkkänen (2008).

I thus adopt a constructivist approach to argument structure, according to which argument structure is independent of (i.e., not projected by) the verb, such that the event/thematic domain is composed of a basic tripartite structure Voice-v-$\sqrt{}$, introducing the external argument, an event (or state) variable, and a category-neutral lexical root respectively (Hale & Keyser 1993; Kratzer 1996; Marantz 1997; Pylkkänen 2008):

(8) $[_{\text{VoiceP}} (\text{XP}_{\text{external}}) [_{\text{Voice'}} \text{Voice} [_{vP} v [_{\sqrt{P}} \sqrt{} (\text{YP}_{\text{internal}})]]]]$

Here, I collapse $vP/\sqrt{P}$ into a single category ($v/\sqrt{P}$) where the difference between the two is not at stake. Additional arguments can be introduced into the structure via dedicated applicative heads (Marantz 1997; Pylkkänen 2008). Specifically, I assume that a high applicative head introduces an additional argument above vP, mediating between an individual (viz. the applied argument) and the event introduced by the verb (9a); and a low applicative head is the complement of v, and mediates between two individuals (9b):

(9)   a. $[_{\text{VoiceP}} \text{XP}_{\text{external}} [_{\text{ApplP}} \text{DP} [_{\text{Appl'}} \text{Appl} [_{v/\sqrt{P}} v/\sqrt{} \text{YP}_{\text{internal}}]]]]$
       b. $[_{\text{VoiceP}} \text{XP}_{\text{external}} [_{v/\sqrt{P}} v/\sqrt{} [_{\text{ApplP}} \text{DP} [_{\text{Appl'}} \text{Appl} \text{YP}_{\text{internal}}]]]]$

3.1   Ditransitive constructions

Cross-linguistically, ditransitive constructions come in two (underlying structural) flavours: prepositional datives (*I gave a book to John*), in which the THEME precedes the RECIPIENT, and DOCs (*I gave John a book*), where the reverse order obtains. *Prima facie*, Old Ibero-Romance DDNR ditransitive constructions look like garden-variety examples of prepositional datives (OSp. *Mando VI VI [tenajas] a estos omnes que mandé las viñas* "I leave **six amphorae each** to the men to whom I

have left the vineyards"; HB&SB 1999: 290); i.e., structures where the indirect object is part of a prepositional phrase (whose head in Old Ibero-Romance is lexicalised by *a* < Latin *ad* "to, towards") at the base of the structure:

(10)   [$_{VoiceP}$ *pro* [$_{VP}$ VI VI tenajas [$_{V'}$ mando [$_{PP}$ a [$_{DP}$ estos omnes]]]]]

However, whilst the indirect object RECIPIENT in Old Ibero-Romance DDNR constructions might *look* like a prepositional dative, the structure in (10) fails to produce the correct scopal properties of the distributive relation between the DistShare (*VI VI tenajas*) and the DistKey (*estos omnes*); nor does the RedNum appear as the complement to V – as we would predict given that RedNum originates in this position in all clausal examples (cf. §2) –, but in its specifier.

For the appropriate distributive interpretation (RECIPIENT > THEME), where six *different* groups of amphorae are given to each man,[3] the DistKey needs to scope over the DistShare, an analysis which receives independent theoretical support from existing syntactic accounts of binominal *each* (e.g., Beghelli & Stowell 1997; Stowell 2013; cf. §4). This order suggests the configuration proposed for DOCs, where the RECIPIENT is introduced into the event domain in a projection higher than the THEME (Bruening 2010; Harley & Jung 2015; Torres Morais & Lima-Salles 2016), would be more suitable for Old Ibero-Romance DDNR – at least in ditransitive structures – than that of prepositional datives, in which the THEME c-commands the RECIPIENT. Indeed, the linear order (and surface scope) in all but two of the attested ditransitive examples in my corpus reflects the underlying order of DOCs;[4] that is, the indirect object RECIPIENT precedes the RedNum direct object, which often surfaces sentence-finally (cf. IML. *mando* [$_{DistKey}$ *canonicis Bracarensibus*] [$_{DistShare}$ *duos duos morabitinos*] "I leave **two *maravedís* each** to the canons of Braga"; Barbosa Morujão 2010: 145).

3.2   Extending the analysis

Evidence that a double object(-like) configuration underlies DDNR constructions more generally is supported by the data from DDNR in transitive constructions, despite the difference in the number of arguments. Closer inspection of the transitive constructions in the corpus, all of which bar two come from Old Spanish,

---

3.   Contextually, the only felicitous distributive reading is where a different asset goes to a different person (RECIPIENT > THEME) since it does not make sense for the *same* assets in a will to be distributed to different recipients (*THEME > RECIPIENT), a characterisation that seems appropriate to extend to Old Ibero-Romance DDNR constructions and their scopal properties more generally.

4.   These word order alternations can be reasonably derived in Old Ibero-Romance assuming discourse-driven movement.

shows that these structures fall into one of three descriptive categories: transfer (11a), possession (11b), or the implied transfer of possession (11c):

(11) a. *dedes*   ***tres   tres   omnes*** *buenos*   (OSp.)
       give.2PL three three men   good
       "You will each put forward three upstanding men"   (1275; CORDE)

   b. *Los costados ovieron en luengo **seis seis cobdos.***   (OSp.)
       the sides   had.3PL in length six six cubits
       "The sides measured six cubits each in length."   (1275; CORDE)

   c. *Fazen   de fuero **tres   tres   sueldos.***   (OSp.)
       make.3PL of tax   three three soldos
       "They pay three *soldos* each in tax."   (1293; CORDE)

Despite the absence of an indirect object, I hold that the set of constructions in (11a-c) has a shared interpretation with Old Ibero-Romance DDNR ditransitives: that is, all the expressions involve the intended, or actual, possession of the THEME, either by a POSSESSOR or a (non-expressed) benefactive RECIPIENT/GOAL. Accordingly, the interpretative differences between (11a-c), and in turn ditransitive DDNR, are reducible to whether the possession relation expressed involves a static relation (11b) or directional and dynamic transfer (11a,c) and ditransitive DDNR.

   Taking together the two descriptive categories of transitive 'transfer' (illustrated by (11a) and (11c)), the interpretation of such sentences in fact carries, I argue, an implicit RECIPIENT/GOAL. In my corpus, such examples typically involve payment of the direct object THEME to an unspecified RECIPIENT (e.g. OSp. *3 3 meajas deven los pecheros* "the taxpayers owe [*someone*] three *meajas* each"). In (11a), the implicitly triadic nature of the construction is underscored by the deployment of the usually ditransitive verb *dar* "to give". To account for this reading, it is possible to postulate the presence of an abstract RECIPIENT/GOAL constituent – a DP, presumably, in line with its overt counterpart in ditransitive structures – in the syntactic structure of such expressions.[5]

   This proposal is in fact supported by transitive constructions with possession-only readings, as exemplified by (11b). Whereas 'transfer' expressions involve a benefactive RECIPIENT as the indirect object, in the possession-only expressions it is the subject itself that functions as the POSSESSOR, and the transfer interpretation is unavailable. The differences between these expressions and the 'transfer' structures might suggest at first blush that a unified analysis is inappropriate for (11a–c), and thus, in turn, transitive and ditransitive DDNR. A crucial piece of evidence of the

---

5.   Compare similar (non-DDNR) constructions from the period in which an indirect object RECIPIENT – i.e., the overt counterpart of the predicted covert RECIPIENT – is attested (OPt. *cada casal destes* fazer de foro **a el Rey** *hum ferro d'arado* "each household of these will pay in tax **to the King** a plough", CdPt).

unified nature of the proposed grouping, however, is the small number of transfer constructions in my corpus, as illustrated by (12), in which the benefactive RECIPIENT/GOAL is not the indirect object, as in the other examples in the corpus, but the subject itself:

(12) *Aquestos aguardadores ayan   por gualardon … **quatro quatro***
      these   guards   have.SBJV.3PL as back.payment four   four
      marauedis.   (OSp.)
      maravedís
      "Those guards should receive/be paid four maravedís each."   (C13; CORDE)

That is, (12) shares with the 'implicitly triadic' constructions its interpretation (i.e., directional/dynamic transfer), but its missing argument is not the benefactive RECIPIENT but the *AGENT* of the transfer. Conversely, it does not share the possessive structures' reading (i.e., of static possession), but both constructions' subjects function as POSSESSORS, differing only in that (12) involves an *intended* POSSESSOR (i.e., the GOAL of the dynamic transfer), and there is no argument corresponding to the AGENT of a transfer. Table 1 summarises the thematic roles and their corresponding structural constituents in each of the different types of DDNR:

Table 1. Thematic roles and corresponding structural constituents in Old Ibero-Romance DDNR

	Agent	(Possessed) theme	Possessor/recipient/goal	Examples
Ditransitive	subject	direct object	indirect object	3, 13a
'Transfer' transitive (subject as AGENT)	subject	direct object	(implied in interpretation)	11a, 13b
'Transfer' transitive (subject as GOAL)	n/a	direct object	subject	12, 13c
Static possession	n/a	direct object	subject	11b, 13d

In other words, (12), and by extension, the possessor constructions without transfer, look rather like AGENT-less passive structures (cf. Burzio 1986) of their ditransitive counterparts. As such, I propose the subject in these constructions originates in the structural locus of the RECIPIENT,[6] before raising to fulfil the relevant case

---

6. An anonymous reviewer questions why my proposal essentially posits an indirect object passive analysis for DDNR when such structures are not found in Romance (Sp. *La mujer fue dada un regalo* "the woman was given a gift"). To explain this apparent anomaly, I follow Pineda & Royo (2017: 452) in their analysis of case assignment in Modern Romance transfer (of possession) predicates where the RECIPIENT exhibits a dative/accusative alternation (Catalan *Ell li*DAT/*el*ACC *paga/roba* "he pays/robs him"DAT/ACC"). To account for the observation that these

requirements (i.e., to check/value [+nominative] Case features in SpecTP, or theoretical equivalent). On this analysis, the sentence in (12) thus illustrates the RE-CIPIENT argument surfacing in a derived position, as corroborated in my corpus by examples in which the surface order corresponds to the proposed base-generated analysis (e.g., (5)).

The syntactic configuration for each of the foregoing subtypes of DDNR is as follows:[7]

(13) a. Ditransitive DDNR:
   [$_{VoiceP}$ *pro* [$_{v/\sqrt{P}}$ mando [$_{FP1}$ a meus testamenteiiros$_{DistKey}$ [$_{FP2}$ **3 3 libras**$_{DistShare}$]]]]

   b. Transitive DDNR involving 'implicitly triadic' transfer (subject as AGENT):
   [$_{VoiceP}$ *pro* [$_{v/\sqrt{P}}$ dedes [$_{FP1}$ RECIPIENT$_{DistKey}$ [$_{FP2}$ **3 3 omnes**$_{DistShare}$]]]]

   c. Transitive DDNR involving 'implicitly triadic' transfer (subject as RECIPIENT):
   [$_{VoiceP}$ [$_{v/\sqrt{P}}$ ayan [$_{FP1}$ aquestos aguardadores$_{DistKey}$ [$_{FP2}$ **4 4 marauedis**$_{DistShare}$]]]]

   d. Transitive DDNR involving static possession:
   [$_{VoiceP}$ [$_{v/\sqrt{P}}$ ovieron [$_{FP1}$ las varas$_{DistKey}$ [$_{FP2}$ **6 6 cobdos**$_{DistShare}$]]]]

3.3 A unified analysis

That the surface position of possessive *have*'s POSSESSOR argument is a derived one has theoretical precedent in the Kayne (1993)/Freeze (1992) tradition that assumes *have* to be a version of *v*BE (cf. Harley 1995, 2002) with an incorporated abstract adposition along the lines of (14):

(14)
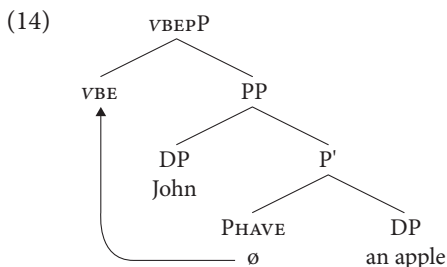
accusative-marked RECIPIENTS can passivise (Catalan *En Joan ha estat pagat/robat* "John has been paid/robbed"), these authors propose that the (non-derived) RECIPIENT is a *structural* accusative, enabling it to passivise like the indirect object of Germanic DOCs. However, the RECIPIENT does not bear *inherent* accusative case, which is assigned – in their analysis and the one I put forward in §3.3 –, by an applicative head to its complement, viz. the THEME.

7. For expository convenience, I do not specify the nature of the projections which merge beneath *v*/√P. In §3.3, I will propose that these layers correspond to ApplP (cf. 9a–b).

On this approach, lexical *have* behaves like an unaccusative verb, with the subject argument (John) merging first in the complement domain of the verb before moving to a higher position to fulfil the nominative Case requirement.

A key appeal of such an approach for the present analysis comes from the implementation of (14) in DOC analyses (Harley 1995, 2002; Harley & Jung 2015) which formalise the intuition that *give*'s meaning is decomposable into the paraphrase "cause to have". Specifically, Harley & Jung (2015) propose that *give* is the surface realisation of P_{HAVE}'s incorporation into a causative light verb (*v*CAUSE) instead of *v*BE. This would straightforwardly unite the DDNR structures discussed up to this point: transfer constructions would involve *v*CAUSE, and possession-only constructions *v*BE, with the subject merging in the specifier of VoiceP in the former, and $P_{HAVE}P$ in the latter. RedNum would merge as the complement of P_{HAVE} and the RECIPIENT/POSSESSOR in its specifier in both cases.

Nonetheless, despite the advantages of the $P_{HAVE}P$ hypothesis,[8] I contend that there are a number of reasons to favour an applicative account – viz. an applicative head introducing the (indirect object) RECIPIENT, and relating it to the direct object THEME – over the $P_{HAVE}P$ hypothesis. Firstly, it has been proposed that an applicative analysis does better than $P_{HAVE}P$ at accounting for the properties of DOCs (Pylkkänen 2008; Bruening 2010). Secondly, the applicative analysis not only captures the asymmetric relation between the internal argument and the indirect object, but also the additional nature of the 'implied' indirect object (which, I have proposed, may be covertly present) in DDNR 'implicitly triadic' agentive transitives involving dynamic/directional transfer. The $P_{HAVE}P$ account, by contrast, does not provide a comparably elegant mechanism for privileging the direct object THEME over the 'additional' indirect object RECIPIENT.[9]

Thirdly, an applicative structure is proposed to underlie a range of constructions in modern Ibero-Romance (Cuervo 2003; Demonte 1995; Pineda 2013; Torres Morais & Lima-Salles 2016) which show similarities with DDNR constructions, including canonical transfer (15a), benefactive/malefactives (15b), and external possession (15c) constructions:

(15) a. *Juan (le)   dio   el   libro a María.*   (ModSp.)
          Juan to.her gave the book to María
          "Juan gave the book to María."

---

8. Cf. Corr (2019) for further details of the relative merits of each approach.

9. Relatedly, derived-subject RECIPIENTS in transfer constructions remain unproblematic under the applicative treatment, since a low Appl analysis has already been independently offered for comparable structures in modern Romance where an accusative-marked RECIPIENT can – unlike its dative counterpart – undergo passivisation (cf. fn. 6).

b. *O João roubou a namorada ao irmão/ roubou-lhe a*
the João robbed the girlfriend to.the brother robbed=to.him the
namorada. (ModPg.)
girlfriend
"João stole his brother's girlfriend."

c. *Juan le lavó la ropa al niño.* (ModSp.)
Juan to.him=washed the clothes to.the child
"Juan washed the child's clothes."

The above-cited authors analyse these constructions as involving a low Appl that introduces a possessor argument denoting a static or dynamic possessor relation between two items:

(16) $[_{\text{VoiceP}} \text{XP}_{\text{external}} [_{v/\sqrt{P}} v/\sqrt{} [_{\text{ApplP}} \text{DP} [_{\text{Appl'}} \text{Appl } \text{YP}_{\text{internal}}]]]]]$

The relation instantiated in each instance is either a dynamic and directional transfer of possession via the asymmetric relationship between the direct object and a RECIPIENT applicative *TO* (15a), or a SOURCE applicative *FROM* (15b); or a static relationship of possession between the possessed direct object, and the possessor argument introduced by a low applicative *AT* (15c). In other words, the structure in (16) already captures the range of structures and interpretations attested in the various types of Old Ibero-Romance DDNR. Furthermore, the exceptional token in §2 of RedNum in a PP-argument (7) provides empirical substantiation of the covert applicative head, which can, on this applicative account, be straightforwardly understood to be lexicalised by the preposition *com*.

   Adopting a (low) applicative analysis for DDNR thus affords our approach independent empirical and theoretical support without sacrificing the assumptions made for the $P_{\text{HAVE}}P$ analysis. Further, it does not restrict DDNR to a possessor(-like) construction, but opens up the possibility for DDNR to encode a wider range of meanings; moreover, it situates the phenomenon within a more general theory of event structure. On this alternative hypothesis, we can simply assume that (i) the (incorporated) adposition in the $P_{\text{HAVE}}P$ analysis is Appl, and (ii) the functional structure labelled as FP1 and FP2 in (13a–d) corresponds to ApplP, where Appl selects RedNum as its complement and DistKey its specifier.

4. A distributive analysis

What does this all have to do with distributivity? Recall that we required a configuration that produced the correct scope for Old Ibero-Romance DDNR and which accounted for the subject-object asymmetry observed. The applicative analysis fulfils both these requirements, and the relational role played by Appl produces the

required dyadic relation between two individuals – here, the RedNum, and the target DistKey over which it distributes its denotation – enabling the distributive interpretation to be read off the clausal syntax. Promisingly, the structural configuration I have proposed mirrors the analyses hypothesised for English binominal *each* by Beghelli & Stowell (1997) and Stowell (2013), suggesting that the present analysis is on the right track. Specifically, these authors postulate a dedicated distributive functional structure involving two contiguous projections, originally labelled DistP and ShareP (Beghelli & Stowell 1997) and recast in Stowell (2013) as a complex DistP shell (composed of Dist1 and Dist2, parallel to DistP/ShareP). In these proposals, an *each*-QP (i.e., the DistKey) merges in the specifier of the higher layer, and the DistShare merges in the lower layer.

Their account diverges from my own, however, in that their distributive layer is (optionally) projected above the event domain. I argue that it is preferable to retain DDNR's analysis within the event domain, for three chief motivations: firstly, the present analysis is reliant on various types of incorporation into $v°$; secondly, keeping our account of distributivity within the event domain is a way of acknowledging the cross-linguistic observation that distributivity is associated with events (cf. §1.1); thirdly, because it enables the present proposal to account for a wider range of meaning (e.g., possession, transfer, etc.) beyond distributivity.

Indeed, the evidence from DDNR in unaccusative structures not only highlights the eventive nature of distributivity, but also provides further motivation for revising our theoretical proposal in favour of an applicative analysis. The three examples of such constructions in the corpus are notable in that all three involve change-of-location verbs with directional PPs plus a benefactive argument (cf. (6)). Fortuitously, they all describe the same event, viz. the Biblical narrative of the animals entering Noah's ark "two at a time". That each example involves an internal argument RedNum ("two two animals") as well as a benefactive argument (Noah) supports the basic intuition of the analysis I have defended thus far (i.e., that DDNR involves an underlying object plus a benefactive RECIPIENT/POSSESSOR corresponding structurally to either the subject or indirect object). However, since Noah is the beneficiary of the animals entering the ark (i.e., the event), not of the animals themselves (the individuals), we can conclude that, even though a benefactive argument is part of the construction, it is *not* the target of the DistShare. That is, unlike the other DDNR examples (in which RedNum is an argument of the verb), all of which involve a distributive relation between individuals, unaccusative DDNR constructions involve distribution over an event.

On an applicative account, the difference between the event-denoting unaccusative examples and the participant-denoting (di)transitive examples comes 'for free', as this theory already has a way of accounting for the above-observed variation, viz. the alternation between 'high' and 'low' applicatives (cf. §3). Thus

the only difference between the event-denoting unaccusative examples and the participant-denoting (di)transitive examples is that the former would involve a high ApplP above *v*P/√P, rather than the low ApplP we have already assumed for the latter. This would entail the (high) applicative head taking the event as its complement (enabling the benefactive argument to benefit from the event itself), rather than RedNum directly; i.e., RedNum would stay within the complement domain of the applicative, preserving the close relation to the benefactive argument, but sufficiently distant from it to avoid establishing a (distributive) relation with the applied argument. That a distributive relationship can be established between two arguments only when these are base-generated in a sufficiently local configuration – specifically, within *v*P – suggests a phase-based explanation for the empirical facts of Old Ibero-Romance DDNR (i.e., it excludes arguments merged outside the *v*P phase, such as high applicative arguments or the external argument).

4.1 Accounting for Old Ibero-Romance RedNum

Finally, we must account for the double duty of RedNum's cardinal numeral, which simultaneously indicates the DistShare DP's quantity and its status as the DistShare. In line with the constructivist approach taken in this paper, and the characterisation of Old (Ibero-)Romance as a(n 'increasingly') configurational language (cf. Ledgeway 2012), I interpret the numeral reduplication as a requirement that DistShare's distributive property be encoded syntactically within its own extended functional structure. To account for its cardinal quantity, I straightforwardly assume a Num[eral]P within the RedNum's extended nominal structure whose specifier is occupied by the cardinal (an XP, following, e.g., Giusti 1997):

(17)   [DP … [NumP **dos**… [NP medidas …
       "two omers"   (Old/Modern Ibero-Romance)

In the spirit of a cartographic approach, I assume that distributivity (or some proxy thereof) is encoded in a dedicated layer within the extended nominal structure; specifically, I assume that the cardinal XP must move to the specifier of this layer to check/value the distributive feature associated with this FP (e.g., through Spec-Head agreement). Independent support for the encoding of distributivity in the nominal functional structure is found in Ouwayda (2014), which argues on the basis of Lebanese Arabic data that a distributive reading is encoded higher in the extended nominal structure than a collective one. Thus I propose that the reduplicated numeral (i.e., the morphosyntactic marker of distributivity in DDNR) corresponds to a non-deleted copy of the cardinal XP as follows (note that the distributive layer is represented here by 'FP'):

(18) a. [DP … [FP **dos** … [NumP ~~dos~~ … [NP medidas …
           "two omers each"   (Old/Modern Ibero-Romance)
      b. [DP … [FP **dos** … [NumP **dos** … [NP medidas …
           "two omers each"   (Old/*Modern Ibero-Romance)

In non-RedNum languages (e.g., Modern Ibero-Romance, English), the lower copy is deleted, necessitating an overt lexical strategy (e.g., Sp. *cada uno/a* "each [one]") at the clausal level for disambiguation between the various possible plural readings. In RedNum languages like Old Ibero-Romance, however, overt spell out of both copies is permitted, thus eliminating the semantic ambiguity at the clausal level and, consequently, the need for a compensatory mechanism, as the distributive reading is already guaranteed.

5.   Conclusion

In addition to contributing new empirical evidence of the phenomenon, the present paper offers, for the first time, a formal analysis of Old Ibero-Romance DDNR. This analysis accounts for the nature of DDNR's surface morphosyntactic expression – viz. a requirement that distributivity be marked in the nominal structure, here via a non-deleted copy of the numeral – and for the sentence-level locus of its semantic effect, i.e., via an independently-motivated (applicative) theory of how argument/event structure interacts with syntactic structure. As well as unifying the range of DDNR constructions where RedNum occurs as a verbal argument, my proposal situates the phenomenon empirically within a wider typology of Ibero-Romance clausal constructions, and, theoretically, within a constructivist approach to the encoding of clausal meaning on which (distributive, in this case) semantics and syntax are tightly connected such that the latter determines the former.

References

Balusu, Rahul. 2005. Distributive reduplication in Telugu. In Christopher Davis, Amy Rose Deal & Youri Zabbal (eds.), *Proceedings of the thirty-six annual meeting of the North East Linguistic Society (NELS 36)*, 39–53. Amherst, MA: GLSA.

Beghelli, Filippo & Tim Stowell. 1997. Distributivity and negation: The syntax of each and every. In Anna Szabolsci (ed.), *Ways of scope taking*, 71–107. Dordrecht: Springer. https://doi.org/10.1007/978-94-011-5814-5_3

Bruening, Benjamin. 2010. Double object constructions disguised as prepositional datives. *Linguistic Inquiry* 41(2). 287–305. https://doi.org/10.1162/ling.2010.41.2.287

Burzio, Luigi. 1986. *Italian syntax: A government-binding approach*. Dordrecht: Springer. https://doi.org/10.1007/978-94-009-4522-7

Cable, Seth. 2014. Distributive numerals and distance distributivity in Tlingit (and beyond). *Language* 90. 562–606. https://doi.org/10.1353/lan.2014.0055

Choe, Jae-Woong. 1987. Anti-quantifiers and a theory of distributivity. Amherst, MA: University of Massachusetts Amherst dissertation.

Corr, Alice. 2018. Old Ibero-Romance reduplicated numerals and distance distributivity. *Workshop on (Co-)Distributivity 2018*, Université Paris 8, France, 22–23 February.

Corr, Alice. 2019. Distributive reduplication in Old Spanish numerals. Talk given at Department of Spanish and Portuguese, University of California, Los Angeles. 12 March.

Cuervo, María Cristina. 2003. Datives at large. Cambridge, MA: MIT dissertation.

Demonte, Violeta. 1995. Dative alternation in Spanish. *Probus* 7(1). 5–30. https://doi.org/10.1515/prbs.1995.7.1.5

Donazzan, Marta & Ana Lúcia Müller. 2015. Reduplicated numerals as pluractionals: Distributivity as a window to the individuation of events. *Journal of Portuguese Linguistics* 14. 95–127. https://doi.org/10.5334/jpl.59

Freeze, Ray. 1992. Existentials and other locatives. *Language* 68. 553–95. https://doi.org/10.2307/415794

Gil, David. 1988. Georgian reduplication and the domain of distributivity. *Linguistics* 26. 1039–1066. https://doi.org/10.1515/ling.1988.26.6.1039

Gil, David. 2005. Distributive numerals. In Matthew Dryer & Martin Haspelmath (eds.), *The world atlas of language structures*, 226–229. Oxford: Oxford University Press.

Giusti, Giuliana. 1997. The categorial status of determiners. In Liliane Haegeman (ed.), *The new comparative syntax*, 95–123. London: Longman.

Hale, Kenneth & Samuel Jay Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In Kenneth Hale & Samuel Jay Keyser (eds.), *The view from Building 20: Essays in honour of Sylvain Bromberger*, 53–109. Cambridge, MA: MIT Press.

Harley, Heidi. 1995. Subjects, events, and licensing. Cambridge, MA: MIT dissertation.

Harley, Heidi. 2002. Possession and the double object construction. *Yearbook of Linguistic Variation* 2. 29–68. https://doi.org/10.1075/livy.2.04har

Harley, Heidi & Hyun Kyoung Jung. 2015. In support of the P$_{HAVE}$ analysis of the double object construction. *Linguistic inquiry* 46(4). 703–730. https://doi.org/10.1162/LING_a_00198

Horcajada Diezma, Baustista & Pedro Sánchez-Prieto Borja. 1999. La reduplicación distributiva del numeral y el arabismo morfosintáctico en el romance hispánico medieval. *Zeitschrift für romanische Philologie* 115. 280–299. https://doi.org/10.1515/zrph.1999.115.2.280

Kayne, Richard S. 1993. Toward a modular theory of auxiliary selection, *Studia Linguistica* 47. 3–31. https://doi.org/10.1111/j.1467-9582.1993.tb00837.x

Kratzer, Angelika. 1996. Severing the external argument from its verb. In Johan Rooryck & Laurie Zaring (eds.), *Phrase structure and the lexicon*, 109–137. Dordrecht: Springer. https://doi.org/10.1007/978-94-015-8617-7_5

Ledgeway, Adam. 2012. *From Latin to Romance: Morphosyntactic typology and change*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199584376.001.0001

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania working papers in linguistics* 4(2). 14. 201–225.

Morujão, Maria do Rosário Barbosa. 2010. *Testamenti ecclesiae portugaliae, 1071–1325*. Lisbon: Universidade Católica Portuguesa.

Myler, Neil. 2016. *Building and interpreting possession sentences*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262034913.001.0001

Ouwayda, Sarah. 2014. Where number lies: Plural marking, numerals, and the collective-distributive distinction. Los Angeles, CA: University of Southern California dissertation.

Pineda, Anna. 2013. Double object constructions and dative/accusative alternations in Spanish and Catalan: A unified account. *Borealis*. 2(1). 57–115. https://doi.org/10.7557/1.2.1.2524

Pineda, Anna & Carles Royo. 2017. Differential indirect object marking in Romance (and how to get rid of it). *Revue Roumaine de Linguistique* 62(4). 445–462.

Postal, Paul M. 1974. *On raising*. Cambridge, MA: MIT Press.

Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge, MA: MIT press. https://doi.org/10.7551/mitpress/9780262162548.001.0001

Ribeiro, Joao Pedro. 1798. *Observações históricas e críticas*. Lisbon: Academia Real das Sciencias de Lisboa.

Safir, Ken & Tim Stowell. 1988. Binominal each. In J. Blevins & J. Carter (eds.), *Proceedings of North East Linguistics Society (NELS)* 18, 426–450. Amherst, MA: GLSA.

Sánchez-Prieto Borja, Pedro & Bautista Horcajada Diezma. 1994. La reduplicación del numeral en textos medievales: ¿una estructura distributiva en castellano antiguo? *Zeitschrift für romanische Philologie* 110. 146–152.

Silva Villar, Luis. 2015. Origen de las estrategias distributivas por reduplicación en espanol y portugués antiguos. *European Scientific Journal*, Special Edition Volume 1, 191–199.

Stowell, Tim. 2013. Binominal each: A DP that may not be. In Kook-Hee Gil, Stephen Harlow & George Tsoulas (eds.), *Strategies of quantification*, 260–294. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199692439.003.0012

Torres Morais, Maria Aparecida & Heloisa Maria Lima-Salles. 2016. The external possessor construction in European Portuguese and Brazilian Portuguese. In Mary A. Kato & Francisco Ordóñez (eds), *The morphosyntax of Portuguese and Spanish in Latin America*, 204–235. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190465889.003.0009

Wright, Roger. 2010. Bilingualism and diglossia in medieval Iberian (350–1350). In Fernando Cabo Aseguinolaza, Anxo Abuín González & César Domínguez (eds.), *A comparative history of literatures in the Iberian Peninsula*, 333–350. Amsterdam: John Benjamins. https://doi.org/10.1075/chlel.xxiv.16wri

Zimmermann, Malte. 2002. *Boys buying two sausages each: On the syntax and semantics of distance-distributivity*. Utrecht: Netherlands Graduate School of Linguistics.

CHAPTER 11

Value and cardinality in the evaluation of bare singulars in Brazilian Portuguese

Suzi Lima and Cristiane Oliveira

University of Toronto / Massey University

Previous studies have explored two semantic features – cardinality and volume – to shown that bare singulars such as *banana* in *Eu comi banana* (I ate banana) allows a count and a mass interpretation in Brazilian Portuguese. In this chapter we explore whether other features such as value impact the interpretation of such nouns. The results of our priming and judgment studies indicate that cardinality is grammatically more relevant than value when interpreting Brazilian Portuguese bare singulars. However, we found a significant increase in reaction times when value was primed early. Therefore, value plays a role in the processing of bare singulars.

Keywords: bare singulars, Brazilian Portuguese, priming, semantics

1.   Introduction

Brazilian Portuguese, as other Romance languages such as French, Italian and Spanish, is a language that grammatically distinguishes count from mass nouns (cf. Paraguassu-Martins & Müller (2007) and references therein). Two well-known tests to show this distinction is the distribution of plural morphemes and numerals: only count nouns (such as *cachorro* "dog" and *cadeira* "chair") can be pluralized and directly combined with numerals (*três cachorros* "three dogs"; *três cadeiras* "three chairs"). On the other hand, mass nouns such as *farinha* "flour" and *água* "water" can neither be pluralized nor directly combined with numerals (*três carnes "three meat"/ *três quilos de carne* "three kilos of meat").[1]

---

1.   Except in cases of coercion (universal packager, cf. Gleason 1965; Pelletier 1979; Doetjes 1997; and Wiese & Maling 2005; for experimental studies of coercion in Brazilian Portuguese see Beviláqua, Lima & Pires de Oliveira 2016 and Lima 2019). The universal packager is an operation that refers to clearly individuated and standardized portions of a substance. For example, in a

As opposed to other Romance languages, Brazilian Portuguese is characterized by the high frequency of bare singulars in argument positions. The term 'bare singulars' will be used in this chapter to refer to count nouns that occur in argument position without determiners or number morphology; that is, count nouns with a mass syntax. In Brazilian Portuguese, bare singulars occur in generic and episodic scenarios:

(1)   *Cachorro late*   (generic)
       dog   bark
       "Dogs bark."

(2)   *João leu   livro*   (episodic)
       João read book
       "João read (a/the/some) books."

<div align="right">Pires de Oliveira & Mendes de Souza 2013: 34</div>

Bare singulars in Brazilian Portuguese are productive in object position,[2] with no restrictions on the class of verbs they can be combined with, differently from other Romance languages where bare singulars are either unacceptable or marginally accepted (cf. Oggiani (2011) for Uruguayan Spanish).

Previous studies on the interpretation of bare singulars have explored whether the absence of number morphology (mass syntax) would necessarily trigger a mass interpretation or whether the lexical semantic features (of denoting objects) would necessarily trigger a count interpretation, regardless of their syntactic properties. Studies have shown that both interpretations are possible, depending on the context. That is, one may use a bare singular to refer to a cardinality of Xs ((3), count interpretation) as well as to refer to a big X/ big portion of X ((4), mass interpretation):

(3)   *João comprou muito   livro hoje*
       João bought   many.much book today
       "João bought many books today"

---

restaurant one may say "There are three coffees on the table" in order to refer to three cups of coffee; the same speaker won't combine a numeral directly to a substance-denoting noun (three coffees) if this person is referring to three drops of coffee on the floor. As such, coercion is restricted to marked scenarios.

2.   While there is consensus in the literature that bare singulars are productive in the object position regardless of the type of predicate (cf., Pires de Oliveira 2014 and literature cited therein), there is some debate on the contexts and predicates that license bare singulars in the subject position (cf. Menuzzi, Figueiredo Silva & Doetjes 2015; Santana & Grolla 2018, for more details).

(4)   Context: João is travelling and he has thick/heavy books on his hands:
   Quanto   livro   você acha   que pode carregar!?
   how-much-SG book-SG you think that can   to carry
   "What quantity of book can you carry?!"
   É muito   livro   pra você levar!
   Is much-SG   book-SG for you to carry.
   "That quantity of book(s) is too much for you to carry."

   Pires de Oliveira & Rothstein (2011: 2172)

Previous studies have shown that in neutral contexts, the preferred interpretation of bare singulars is a count interpretation (Lima 2014; Lima & Gomes 2016; Beviláqua, Lima & Pires de Oliveira 2016). For example, Lima (2014) describes the results of a quantity judgment task where 38 Brazilian Portuguese speakers had to answer *Quem tem mais X?* "Who has more X?" – where X could be a bare singular (*galinha* "chicken"), a substance mass noun (*água* "water") or a bare plural (*galinhas* "chickens") – while being presented with two pictures: a picture that included a cardinality of Xs and a picture that included a big X/big portion of X. The participants favored the cardinal response when the questions included a bare singular or a bare plural (86% and 100% of cardinal responses, respectively) and they disfavored the cardinal response when the question included a mass noun (21% of cardinal responses).[3]

While in neutral contexts, as discussed above, the favored interpretation for bare singulars seems to be a cardinal interpretation, some studies have shown that contextual factors may impact the results of quantity judgment tasks. Beviláqua & Pires de Oliveira (2014) have shown that the mass interpretation of bare singulars can be favored by the context. A sample of their experimental stimuli is presented in Figure 1. Beviláqua & Pires de Oliveira created experimental items where the volume interpretation was favored. For example, instead of being presented with the question *quem tem mais bola?* "Who has more ball?" in a neutral context, the participants were presented with the question *quem tem mais bola **para encher o cesto**?* "Who has more ball *to fill the basket*?" in which case the size/volume of the objects is more relevant than their cardinality. In their task, participants could choose among four possible answers: (a) Maria (cardinal response); (b) Joana (volume response); (c) Maria or Joana; (d) neither Maria nor Joana. Beviláqua & Pires de Oliveira (2014: 90) observed that a volume response was favored (60% of volume responses (option b); 20% of cardinal responses (option a) and 20% of cardinal and volume responses (option c)).

---

3.   The results of this task were replicated in Lima & Gomes (2016). In a truth value judgment task with 22 speakers where the participants were exposed to explicit comparison (A has more X than B). The cardinal response was favored for sentences that included bare singulars and bare plurals (99% and 100%, respectively) and disfavored for mass nouns (21% of number responses). See also Beviláqua et al. (2016) for similar results.
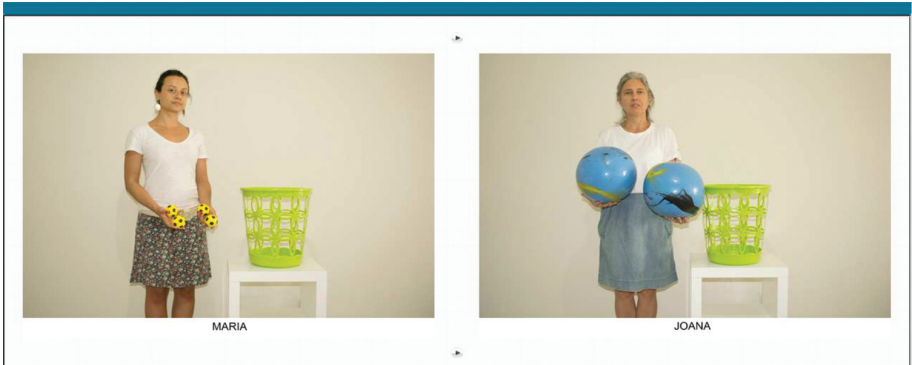
Figure 1. Context: *Joana e Maria querem encher o cesto* ("Joana and Maria want to fill the basket"); Question: *Quem tem mais bola para encher o cesto?* ("Who has more ball(s) to fill the basket?") (Beviláqua & Pires de Oliveira 2014)

In sum, results of tasks exploring the count and the mass interpretation of bare singulars have shown that the interpretation of bare singulars is strongly impacted by its lexical semantic features in neutral contexts.[4] That is, we have seen that despite the fact that bare singulars allow count and mass interpretations, the count interpretation is preferred for bare singulars in neutral contexts. While both the features cardinality (cardinality of X) and volume (big X/big portion of X) affect the interpretation of bare singulars, cardinality overrules volume in neutral contexts. Therefore, the interpretation is not solely guided by syntax.

Studies in languages such as English have shown that semantic features such as functionality, heterogeneity or value impact the interpretation of nouns (Giralt & Bloom 2000; Grimm & Levin 2012; Srinivasan et al. 2013). As such, in this chapter, we want to explore whether there are other relevant dimensions of comparison – other than cardinality – in the interpretation of bare singulars in Brazilian Portuguese. In § 2 we present an overview of the literature that explores dimensions of comparison other than cardinality in English. In § 3 we present the results of two experimental studies.

---

4. It is interesting to note that in languages where bare singulars are not productive, we see a different pattern. In quantity judgments tasks in English with flexible nouns (count nouns that can be in a count and mass syntax such as *stones/stone, chocolates/chocolate*), Barner & Snedeker (2005) observed that the bare form is more likely to be associated with a mass interpretation than with a count interpretation.

2.   Dimensions of comparison beyond cardinality

Much literature has shown that features such as functionality and heterogeneity impact the interpretation of nouns. Acquisition studies with English speaking children have shown that while children may count parts of objects as wholes (for example, they may count a broken fork as two forks, cf. Shipley & Shepperson 1990; Gutheil et al. 2004; Srinivasan et al. 2013, among others) this pattern is not maintained if children know the function of the parts. For instance, when presented with a wheel of a bicycle detached from its frame, children won't count these two parts as two bicycles if they know the name of the parts and their functionality (Giralt & Bloom 2000; Srinivasan et al. 2013). Grimm & Levin (2012) have shown the impact of features such as functionality and heterogeneity on the interpretation of object-mass nouns such as *furniture*.[5] In quantity judgment tasks, participants evaluated whether someone with a small cardinality of valuable/more elaborate jewelry had more jewelry than someone with a higher cardinality of less valuable jewelry (5). As such, this illustrates the impact of the feature value in the interpretation of nouns:

> (5)   Sample context: Two women are at a gala event. Woman A is wearing two gold bracelets, a diamond tiara, and a ruby and emerald necklace (4 items). Woman B is wearing three gold rings, a pearl necklace and a silver bracelet (5 items). Who has more jewelry?   (Grimm & Levin 2012)

Two important observations may be gathered from the English data discussed so far. First, the literature has shown that in quantity judgment tasks (cf. Barner & Snedeker 2005), object-mass nouns in English favor a cardinal interpretation – as we saw for bare singulars in Brazilian Portuguese. Second, the literature has shown that the interpretation of object-mass nouns is subject to features such as heterogeneity and functionality in English. From the literature on Brazilian Portuguese (as discussed in § 1), we have learned that cardinality is more relevant than volume in neutral contexts. What the literature on Brazilian Portuguese bare singulars hasn't shown yet is whether there are other features that may impact the interpretation of bare singulars, as observed for object-mass nouns in English. Thus, in this chapter, we want to investigate whether value can impact the interpretation of bare singulars in Brazilian Portuguese. In order to investigate that, we present the results of two studies: one task where we manipulated two types of contexts (the feature value was relevant in one context and not critical in the other); second, we present the results of a priming task where we manipulated the features value and cardinality.

---

5.   In this chapter we take object mass nouns to be nouns that denote objects but have the same distribution as of mass nouns. That is, they can neither be pluralized nor directly combined with numerals (**three furnitures; three pieces of furniture*) (cf. Schwarzschild 2011, among others).

3.   Studies

3.1   Study 1: Truth value judgment task

3.1.1   *Materials and methods*

In a truth value judgment task (Google forms) 40 Brazilian Portuguese participants were presented with short narratives followed by a sentence that they had to evaluate as true or false, given the scenario. A total of 20 short narratives were created and divided in two lists. Each participant was exposed to 10 short narratives and 12 fillers unrelated to the manipulation. The design was between subjects: 20 participants were exposed to 10 short narratives of what we will call here the 'neutral contexts', that is, contexts where the feature value was not prominent (6a–6b). The other 20 participants were exposed to 10 short narratives of what we will call here the 'value contexts' where this feature was prominent (7a–7b). Each context featured two establishments or two people (A and B). Either person/establishment A had a larger cardinality of objects than the person/establishment B or person/establishment A had more valuable items than the person/establishment B.

'Neutral Context' (CVV): Cardinality (DP_1), Value (DP_2); Target sentence's first DP: Value (DP_2)

(6)   a.   *Maria quer comprar bonecas para distribuir como brinde em uma feira. Ela foi a duas lojas. Na **loja Praia** ela encontrou cinco bonecas de plástico sem roupa. Na **loja Mar** ela encontrou três bonecas Barbie recém-lançadas.*
*Na **loja Mar** tem mais boneca que na **loja Praia**.*
Verdadeiro   Falso
"Maria wants to buy dolls to distribute at a fair. She went to two stores. In the **Beach store** she found five plastic dolls without clothes. In the **Sea store** she found three newly released Barbie dolls.
The **Sea store** has more doll(s) than the **Beach store**.
True   False"

'Neutral Context' (VCV): Value (DP_1), Cardinality (DP_2); Target sentence's first DP: Value (DP_1)

(6)   b.   *Pedro está olhando vitrines de lojas de beleza. Na **loja Sol** ele encontrou três perfumes Dior. Na **loja Lua** ele encontrou cinco perfumes Boticário.*
*A **loja Sol** tem mais perfume que **a loja Lua**.*
Verdadeiro   Falso
"Pedro is shopping for beauty products. In the **Sun store** he found three Dior perfumes. In the **Lua store** he found five Boticário perfumes.
The **Sun store** has more perfume(s) than the **Lua store**.
True   False"

'Value Context' (CVV): Cardinality (DP$_1$), Value (DP$_2$); Target sentence's first DP: Value (DP$_2$)

(7) a. *A Maria quer presentear suas filhas com bonecas esse ano. Ela realmente quer fazer algo especial para suas filhas já que economizou dinheiro pra isso. Ela foi a duas lojas. Na **loja Praia** ela encontrou cinco bonecas de plástico sem roupa. Na **loja Mar** ela encontrou três bonecas Barbie recém-lançadas. Na **loja Mar** tem mais boneca que na **loja Praia**.*
Verdadeiro   Falso
"Maria wants to give dolls to her daughters this year. She really wants to do something special for her daughters since she saved money for it. She went to two stores. In the **Beach store** she found five plastic dolls without clothes. In the **Sea store** she found three newly released Barbie dolls. The **Sea store** has more doll(s) than in the **Beach store**.
True   False"

'Value Context' (VCV): Value (DP$_1$), Cardinality (DP$_2$); Target sentence's first DP: Value (DP$_1$)

(7) b. *Pedro economizou dinheiro para dar um presente especial para sua mãe. Na **loja Sol** ele encontrou 3 perfumes Dior. Na **loja Lua** ele encontrou 5 perfumes Boticário.*
*A **loja Sol** tem mais perfume que a **loja Lua**.*
Verdadeiro   Falso
"Pedro saved money to give his mother a special gift. In the **Sun store** he found 3 Dior perfumes. In the **Moon store** he found 5 Boticário perfumes. The **Sun store** has more perfume(s) than the **Moon store**.
True   False"

In this study we investigated whether the value feature was a relevant variable in the interpretation of bare singulars. That is, we investigated whether participants would evaluate as true an assertion that someone has more X than Y, in a context where X has more valuable items than Y, but fewer items. Furthermore, we explored whether it makes a difference if the person/establishment that has more valuable items is presented first or second in the context.

3.1.2 *Results*

The results of this study are presented in Table 1.

Table 1. Percentage of acceptance of sentences that favored *value*

Context + Target sentence	Neutral	Value
$V_{DP1}C_{DP2}V_{\text{Target sentence}}$	12%	17%
$C_{DP1}V_{DP2}V_{\text{Target sentence}}$	17%	25%

Overall, the results suggest that a cardinal response was favored, independently of the context. We found a numerical advantage of value responses in scenarios such as (6a) and (7a), where the person/establishment with more valuable items was presented second (DP$_2$) in the context (CVV: 17% [neutral context]; and 25% [value context]). We present next our priming study.

3.2   Study 2: Priming task

3.2.1   *Materials and methods*

Our second study was a response-priming task (PsyScope X B57, MAC OS). 25 Brazilian Portuguese speakers, undergraduate students from the Federal University of Rio de Janeiro, participated in this study. Each participant answered 92 questions (16 target sentences; 16 controls; 60 filler sentences). In this task, the participants had to read a sentence (as in (8)) and press a button once they finished reading. The priming sentence always contrasted two brands. One was more valuable than the other (a norming task was performed with 37 Brazilian Portuguese speakers in order to select the contrast of brands.).[6]

(8)   *Carla coleciona 10 Rolex e Maria 20 Mondaine.*
       "Carla owns10 Rolex and Maria 20 Mondaine."

Then, participants were shown a target sentence that they had to evaluate as true or false. For example, for (8), the target sentence was *Maria tem mais relógio* "Maria has more watch(es)". Four conditions were manipulated, as illustrated in Table 2.

Given the design of the study (within subjects), all participants were exposed to all conditions. We explored two research questions: first, does the semantic feature of the first DP in the priming sentence affect the evaluation of the target sentence? Second, if it does, will we find different response patterns depending on the semantic features (value or cardinality) of the first DP in the priming sentences?

---

6.   37 Brazilian Portuguese speakers participated in a norming task. Based on a likert scale (scale: 5) participants were presented with a sentence that explicitly said that a particular brand was more valuable than another (*Um relógio Rolex é mais caro que um Mondaine* "A Rolex watch is more expensive than a Mondaine [a not so expensive brand of watches in Brazil]"). The participants had to evaluate whether they agree or disagree with the sentences. We used the results of this norming task to choose the brands to be used in the offline and online studies presented in this chapter.

Table 2. Conditions (priming task)

Condition	Example
Condition VCC: priming [+value] DP1 [+cardinality] DP2	Carla owns10 Rolex and Maria 20 Mondaine. Target sentence : *Maria tem mais **relógio*** "Mary has more watch(es)" [+cardinality].
Condition CVC: priming [+cardinality] DP1 [+value] DP2	Maria owns 20 Mondaine and Carla 10 Rolex. Target sentence : *Maria tem mais relógio* "Mary has more watch(es)" [+cardinality]
Condition VCV: priming [+value] DP1 [+cardinality] DP2	Carla owns 10 Rolex and Maria 20 Mondaine. Target sentence : *Carla tem mais relógio* "Carla has more watch(es)" [+value]
Condition CVV: priming [+cardinality] DP1 [+value] DP2	Maria owns 20 Mondaine and Carla 10 Rolex. Target sentence : *Carla tem mais relógio* "Carla has more watch(es)" [+value]

3.2.2 *Results*

The first observation we make is that, as in our first study, cardinality is preferred over value for most conditions in the target sentence responses, as presented in Table 3.

Table 3. Percentage of preference for the 'cardinality' answer per condition (critical items)

Condition (code)	Conditions (descriptions)	Cardinality answer
VCC	First DP: Value; Second DP: Cardinality Target sentence: Cardinality	80%
VCV	First DP: Value; Second DP: Cardinality Target sentence: Value	74%
CVC	First DP: Cardinality; Second DP: Value Target sentence: Cardinality	76%
CVV	First DP: Cardinality; Second DP: Value Target sentence: Value	59%

Subjects favored the cardinality answer for all conditions (74–80%) except for the condition CVV (59%). The condition type affected the answers of the target sentence (X-squared = 10.382, *df* = 3, p-value = 0.01558*). Table 4 presents the mean of reaction time in milliseconds.

The results we obtained suggest that cardinality interpretations of target sentences are significantly faster when priming from the first DP ([CVC] vs [VCC]

Table 4. Reaction times (critical items) – within subject (t-test)

Condition (code)	Mean reaction times in milliseconds	Cardinality answer
VCC	1286	80%
VCV	1603	74%
CVC	1033	76%
CVV	1662	59%

$t(19) = 3.13$ $p < 0.0055$ **). That is, between CVC and VCC, cardinality interpretations are more costly when the first DP is characterized by the value feature. Another observation that we would like to highlight is that the order of priming has no significant effect on the interpretation of value target sentences [VCV] vs [+CVV] $t(19) = 0.61$ $p < 0.5522$. Overall, CVV answers seem to be at chance (59% of cardinality answers). This could be an effect of the presence of a value DP as the second DP. The same tendency was observed in our offline study (we had more value answers when value was marked in the second DP).

One could ask whether the significant difference between the conditions CVC and VCC was due to the interaction among the value feature and the early processing of the first DP in the priming sentence or if that difference was due to a simple mismatch between the first DP in the priming sentence and the DP in the target sentence. We can better evaluate these results by analyzing the control items, as presented in the next section.

3.3   Control items

3.3.1   *Materials and methods*
The control items of this experiment were 16 sentences with a similar structure as the critical items, as illustrated in Table 5:

Table 5. Conditions (controls)

Condition NCC: priming [−cardinality] $_{DP1}$ [+cardinality]$_{DP2}$	Carla owns 10 blue plates and Maria 20 red plates. Target sentence: *Maria tem mais prato.* "Maria has more plate(s)." [+cardinality]
Condition CNC: priming [+cardinality] $_{DP1}$ [−cardinality]$_{DP2}$	Maria owns 20 red plates and Carla 10 blue plates. Target sentence: *Maria tem mais prato.* "Maria has more plate(s)." [+cardinality]
Condition NCN: priming [−cardinality] $_{DP1}$ [+cardinality]$_{DP2}$	Carla owns 10 blue plates and Maria 20 red plates. Target sentence: *Carla tem mais prato.* "Carla has more plate(s)." [−cardinality]
Condition CNN: priming [+cardinality] $_{DP1}$ [−cardinality]$_{DP2}$	Maria owns 20 red plates e Carla 10 blue plates. Target sentence: *Carla tem mais prato.* "Carla has more plate(s)." [−cardinality]

As explained below, the motivation of the study with the control items was to check whether the higher reaction times of processing the target sentence (Condition VCC in contrast with CVC, § 3.2) is due to the value feature in DP_1 (Condition VCC, priming sentence) or due to the mismatch between DP_1 in the priming sentence and the DP in the target sentence.

3.3.2   *Results*

The results of this study are presented in Table 6.

Table 6. Percentage of preference for the cardinality answer per condition and reaction times (controls)

Condition (code)	Conditions (description)	Cardinality answer	Mean reaction times in milliseconds
NCC	DP_1:[−Cardinality]; DP_2: [+Cardinality] *Target sentence*: [+Cardinality]	78%	1534
CNC	DP_1:[+Cardinality]; DP_2: [−Cardinality] *Target sentence*: [+Cardinality]	82%	1136
NCN	DP_1:[−Cardinality]; DP_2: [+Cardinality] *Target sentence*: [−Cardinality]	80%	1483
CNN	DP_1:[+Cardinality]; DP_2: [−Cardinality] *Target sentence*: [−Cardinality]	75%	1403

As expected, the participants favored the cardinality answer for all conditions (+ cardinality, 75–82%). No condition effect was found (X-squared = 1.4158, $df = 3$, p-value = 0.7018). Furthermore, the contrast between the conditions NCC and CNC ($t(19) = 1.12\ p < 0.2779$) and NCN × CCN $t((19) = 0.68\ p < 0.5034)$ suggests the absence of effects across different conditions, even when the target sentence presented a cardinality mismatch with DP_1 from the prime sentence. Therefore, the effect found in the priming main study (§ 3.2.) was indeed due to the interaction between value feature and early processing of the first DP (priming sentence) and not due to a mismatch between the first DP in the prime sentence and the DP in the target sentence.

4.   Conclusions

In this chapter we have shown that both value and cardinality seem to be features that can affect the interpretation of bare singulars of Brazilian Portuguese. However, as expected, both experiments reveal that cardinality seems to be grammatically more relevant than value in the interpretation of bare singulars, corroborating previous offline studies (cf. Lima 2014; Lima & Gomes 2016; Beviláqua, Lima & Pires de Oliveira 2016, and literature cited therein).

While cardinality has been shown to be more relevant than value in the interpretation of bare singulars, it is important to note that the value feature also plays a role in the processing of bare singulars even when the cardinality feature is activated; the feature value increases the reaction times when primed early in the sentence.

References

Barner, David & Jesse Snedeker. 2005. Quantity judgments and individuation: Evidence that mass nouns count. *Cognition* 97(1). 41–66. https://doi.org/10.1016/j.cognition.2004.06.009

Beviláqua, Kayron & Roberta Pires de Oliveira. 2014. Brazilian bare phrases and referentiality: Evidences from an experiment. *Revista Letras* 90(2). 253–275.
https://doi.org/10.5380/rel.v90i2.37234

Beviláqua, Kayron, Suzi Lima & Roberta Pires de Oliveira. 2016. Bare Nouns in Brazilian Portuguese: An experimental study on grinding. *Baltic International Yearbook of Cognition, Logic and Communication* 11(1). 2. 1–25. https://doi.org/10.4148/1944-3676.1113

Doetjes, Jenny Sandra. 1997. *Quantifiers and selection: On the distribution of quantifying expressions in French, Dutch and English*. The Hague: Holland Academic Graphics.

Giralt, Nuria, & Paul Bloom. 2000. How special are objects? Children's reasoning about objects, parts, and holes. *Psychological Science* 11(6). 497–501.
https://doi.org/10.1111/1467-9280.00295

Gleason, Henry Allan. 1965. *Linguistics and English grammar*. New York: Holt, Rinehart and Winston.

Grimm, Scott & Beth Levin. 2012. Who has more furniture? An exploration of the bases for comparison. Paper presented at Communication au colloque Massif/comptable en linguistique, philosophie et sciences, Paris, France, December 19–21.

Gutheil, Grant, Paul Bloom, Nohemy Valderrama & Rebecca Freedman. 2004. The role of historical intuitions in children's and adults' naming of artifacts. *Cognition* 91(1). 23–42.
https://doi.org/10.1016/S0010-0277(03)00165-3

Lima, Suzi. 2014. The grammar of individuation and counting. Amherst, MA: University of Massachusetts Amherst dissertation.

Lima, Suzi. 2019. Processing coercion in Brazilian Portuguese: grinding objects and packaging substances. In Katy Carlson, Charles Clifton, Jr. & Janet Dean Fodor (eds.), *Grammatical approaches to language processing – Essays in honor of Lyn Frazier*, 209–224. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-030-01563-3_11

Lima, Suzi Oliveira & Ana Paula Quadros Gomes. 2016. The interpretation of Brazilian Portuguese bare singulars in neutral contexts. *Revista Letras* 93. 193–209.
https://doi.org/10.5380/rel.v93i1.42690

Menuzzi, Sérgio de Moura, Maria Cristina Figueiredo Silva & Jenny Doetjes. 2015. Subject bare singulars in Brazilian Portuguese and information structure. *Journal of Portuguese Linguistics* 14(1). 7–44. https://doi.org/10.5334/jpl.56

Oggiani, Morgas C. 2011. On discourse referential properties of bare singulars in Spanish. Utrecht, Netherlands: Utrecht University Master's thesis.

Paraguassu-Martins, Nize & Ana Müller. 2007. A distinção contável-massivo e a expressão de número no sistema nominal. *Revista Delta* 23. 65–83.
https://doi.org/10.1590/S0102-44502007000300006

Pelletier, Francis Jeffrey. 1979. Non-singular reference: some preliminaries. In Francis Jeffrey Pelletier (ed.), *Mass terms: Some philosophical problems*, 1–14. Dordrecht: Springer.
https://doi.org/10.1007/978-1-4020-4110-5

Pires de Oliveira, Roberta. 2014. Dobras e redobras do singular nu: costurando a semântica através das línguas. *Porto Alegre: EDIPUC-RS*.

Pires de Oliveira, Roberta & Luisandro Mendes De Souza. 2013. 'O singular nu e a comparação: uma proposta de derivação semântica'. *Revista LinguíStica* 9(1). 31–54.

Pires de Oliveira, Roberta & Susan Rothstein. 2011. Bare singular noun phrases are mass in Brazilian Portuguese. *Lingua* 121(15). 2153–2175. https://doi.org/10.1016/j.lingua.2011.09.004

Santana, Raíssa Silva & Elaine Grolla. 2018. The acceptability of subject bare singulars in Brazilian Portuguese. *Revista LinguíStica* 14. 194–214.
https://doi.org/10.31513/linguistica.2018.v14n2a17532

Schwarzschild, Roger. 2011. Stubborn distributivity, multiparticipant nouns and the count/mass distinction. In Suzi Lima, Kevin Mullin & Brian Smith (eds.), *Proceedings of the 39th Annual Meeting of the North East Linguistic Society (NELS)*, Volume 2, 661–678. Amherst, MA: Graduate Linguistics Students Association, University of Massachusetts Amherst.

Shipley, Elizabeth F., & Barbara Shepperson. 1990. Countable entities: Developmental changes. *Cognition* 34(2). 109–136. https://doi.org/10.1016/0010-0277(90)90041-H

Srinivasan, Mahesh, Eleanor Chestnut, Peggy Li & David Barner. 2013. Sortal concepts and pragmatic inference in children's early quantification of objects. *Cognitive psychology* 66(3). 302–326. https://doi.org/10.1016/j.cogpsych.2013.01.003

Wiese, Heike & Joan Maling. 2005. Beers, kaffi, and schnaps: Different grammatical options for restaurant talk coercions in three Germanic languages. *Journal of Germanic Linguistics* 17(1). 1–38. https://doi.org/10.1017/S1470542705000012

CHAPTER 12

Formality by distance in Spanish and Catalan

Gavin Bembridge and Andrew Peters
York University / University of Toronto

Unlike the pragmatic literature, a pronoun's ability to distinguish levels of formality has received relatively little dedicated attention in the syntactic literature. One definition of formality from the politeness literature defines it as social distance (Brown & Levinson 1987). We take the idea that formality can be viewed as distance literally and argue that second-person pronouns can incorporate Harbour's (2016) projection, χ, that encodes spatial semantics. Variation in second-person pronouns results from differences in each language's specific pronominal resources and the social meanings they are assigned.

Keywords: second-person pronouns, formality, Spanish, Catalan, Distributed Morphology, semantics, pragmatics

1. Introduction

The linguistic literature on pronouns has looked look at various properties, such as logophoricity (e.g., Sells 1987), binding (e.g., Chomsky 1981), and pronoun strength (e.g., Cardinaletti & Starke 1999), among many others. Though certain pronouns can distinguish levels of formality, this use has received relatively little attention in the syntactic literature. However, the same is not true of the sociolinguistic and pragmatic literature. At least one definition of formality found in this body of work defines it as social distance (Brown & Levinson 1987). We take this idea at face value and argue that second-person pronouns in Spanish and Catalan can incorporate Harbour's (2016) projection, χ (from Greek *xoros* "space"), that encodes spatial semantics. Specifically, in second-person pronouns, χ returns a property of individuals as either being in the vicinity of the author of a Speech Act or not.

This chapter is organized as follows. Section 2 looks at the distribution and function of second-person pronouns in several varieties of Spanish and in Catalan. Sections 3 and 4 collectively outline the relevant theoretical background on person

and number, and on spatial deixis, respectively. In § 5, we lay out our proposal for the syntax and semantics of second-person pronouns. Section 6 addresses the morphological realization of the syntactic structures associated with these pronouns. Last, § 7 concludes.

2.   Second-person pronouns in Spanish and Catalan

This section presents the data we need to account for in Spanish and Catalan. An account of the second-person formality contrasts should permit syncretisms based on formality and provide for a three-way formality contrast in Salvadoran Spanish and Catalan.

In Standard Peninsular Spanish (SPS), two levels of formality are marked in both singular and plural second-person pronouns. In Table 1, in the singular, the informal and formal pronouns are *tú* and *usted* and *vosotros* and *ustedes* are their informal and formal counterparts in the plural.[1]

Table 1. Standard Peninsular Spanish

	SG	PL
Informal	*tú*	*vosotros*
Formal	*usted*	*ustedes*

Standard Latin American Spanish (SLAS) follows almost the same pattern as SPS in the singular, except for countries (i.e., Argentina, Paraguay, Nicaragua, and Honduras) where the pronoun *vos* replaces *tú* as the informal singular pronoun (Lipski 2007). However, in SLAS there is a syncretism in the plural and *ustedes* is used in both informal and formal contexts. This is illustrated in Table 2.

Table 2. Standard Latin American Spanish

	SG	PL
Informal	*tú / vos*	*ustedes*
Formal	*usted*	

Salvadoran Spanish (SS) departs from SLAS displaying a three-way formality contrast in the singular. That is, in SS *vos* and *tú* coexist whereas in the standard case

---

1.   Unless otherwise stated, the grammatical descriptions of Spanish in this section are drawn from Butt & Benjamin (2000).

only one or the other is used. *Vos* is a familiar address for close social relationships, *tú* is used informally and *usted* is the formal address (Lipski 2007: 277). The facts for the plural follow the SLAS pattern as there is a formality syncretism, although in this case there are three categories instead of two. The relevant data are given in Table 3.

Table 3. Salvadorean Spanish

	SG	PL
Familiar	*vos*	
Informal	*tú*	*ustedes*
Formal	*usted*	

Contemporary Cuban Spanish (CS) differs from the other Latin American varieties as it no longer distinguishes formality morphologically. Lipski (2007: 258) notes that in CS, *tú* is the default pronoun and occurs in all cases, including those in which *usted* would be common in other varieties. The pronoun *ustedes* is used as the plural form as illustrated in Table 4.

Table 4. Cuban Spanish

SG	PL
tú	*ustedes*

Last, Catalan strong pronouns also have a three-way formality distinction. Wheeler, Dols & Yates (1999: 161–162) explain that *tu* is the informal address, *vós* is a polite address, and *vostè* is also polite but more distant than *vós*. In the plural, there is a syncretism and *vosaltres* is used in both the informal and respectful contexts; the pronoun *vostès* is the plural counterpart of *vostè*. These facts are summarized in Table 5.

Table 5. Catalan

	SG	PL
Informal	*tu*	*vosaltres*
Respectful	*vós*	
Formal	*vostè*	*vostès*

In the next section, we discuss the syntax and semantics of person and number followed by a discussion of spatial deixis and its relation to pronouns in Section 4.

3.   On person and number

In the syntactic literature, it has been argued that there is a hierarchical structure to the features that make up pronouns (Noyer 1992; and Harley & Ritter 2002). Harley & Ritter (2002) use a feature hierarchy, like the one described for phonology (see Clements 1985), to account for various person contrasts observed crosslinguistically.[2] In their account, pronominal features are divided between participant and individuation features. For example, second-person pronouns have, at minimum, the participant features [−speaker] and [+addressee]. The framework we adopt builds on the notion that a person paradigm may be composed of hierarchically related features.

Harbour's (2016) ontology of atomic persons consists of *i, u,* and *o* which represent the author, the addressee and any others, respectively. These person atoms can be combined in a semilattice, yielding four combinations: i_o, u_o, o_o, iu_o. The subscript *o* represents zero or more others. From these atoms, a language can distinguish first-person singular or plural exclusive, second person, third person, and first-person inclusive, respectively. Although these persons may mathematically be combined in 15 unique partitions, Harbour notes that only five are attested across the world's languages. These five types of contrasts (e.g., 1st vs. non-1st or non-3rd vs. 3rd etc.) may be represented by two person features, [±author] and [±participant]. These features either add to or subtract from the set of persons a pronoun may represent, in a language-specific order, yielding the five types of person contrasts that Harbour observes crosslinguistically. Relevant for Spanish and (other Romance languages) is Harbour's (2016) Standard Tripartition which makes a distinction between a first, second, and third person without clusivity. Consider (1).

(1)   Standard Tripartition (i.e., 1st vs. 2nd vs. 3rd): [±participant] ≫ [±author]

Number contrasts could be accounted for by Harbour's (2014) proposal wherein a number phrase marks for atomicity, minimality and additivity. For languages with a singular-plural distinction, [±atomic] is the feature that contrasts singleton sets from others. Nothing crucial hinges on the adoption of this feature and other theories of number should, in principle, be compatible with our final proposal.

Next, we show how Harbour's (2016) χ operator combines with these person features to produce a 'characteristic space' of the author of speech.

---

2.   Though the explanatory desirability of morphosyntactic feature geometries has been challenged recently, see Harbour (2011) for relevant discussion.

4.   Spatial deixis, vicinities and pronouns

Harbour (2016: 179) argues that the χ operator can take a set of individuals specified on the head containing person features and return a vicinity or characteristic space of that individual. The denotation of the χ operator is given in (2).

(2) $\chi = \lambda P_{\langle e,\, t \rangle} \cdot \lambda y_{\in De} \cdot y \in \chi(x) \wedge P(x)$
     x, a free variable   (Harbour 2016: 179)

Harbour (2016) notes that χ takes a predicate, P, and provides a free variable satisfying that predicate (i.e., $P(x)$) and this creates $\chi(x)$ which is the vicinity of x. The χ operator also provides a variable, y, over the domain of entities (i.e., D_e) that allows us to specify that an entity is in the vicinity of x. He argues that a deictic system with a three-way partition could also be easily modified to build a proximal-medial-distal contrast using binary [participant] and [author] features. This is demonstrated below in (3).

(3)   a.   Proximal: [+participant, +author]
      b.   Medial: [+participant, −author]
      c.   Distal: [−participant]

The χ operator simply combines with the head that bears these features (π in Harbour's syntax) and returns a vicinity relative to the person described in π.
      Similarly, Bjorkman et al. (2019) argue that χ can appear as a feature in pronouns. Their analysis accounts for third-person pronouns in Heiltsuk (Wakashan, BC, Canada) which has a 6 (or 7)-way distinction for proximity to the speaker, hearer, and visibility. Since (third-person) pronouns in some languages may reference a space relative to a discourse participant, they argue that these pronouns must have a structure including a Harbour-style χ operator. However, they argue for a more simplified view of the composition of person features and derive Harbour's (2016) typology by treating the features, [±author] and [±participant], as first-order predicates whose ordering represent contrastive scope relations.
      This mechanism of positioning spaces relative to discourse participants is central to distinguishing degrees of formality in second-person pronouns in Spanish and Catalan. That is when second-person pronouns include the χ operator the result is a calculation of 'psychological distance' or formality. We more concretely formalize this account in the following section.

5.   The proposal

If second-person pronouns are simply [+participant, −author] as has been sug-
gested widely in the literature (Halle 1997; Harley & Ritter 2002; Harbour 2016,
among many others), then the set of pronouns that these features characterize in
Spanish and in Catalan would be as follows:

(4)   a.   [+participant, −author] = {*vos, tú, usted, vosotros, ustedes*}   [Spanish]
    b.   [+participant, −author] = {*tu, vosaltres, vostè, vostès, vós*}   [Catalan]

However, this feature set does not account for the fact that these pronouns encode
differences in (i) number and (ii) formality. While number is simpler to deal with in
Harbour's (2016) system, there is no intuitive way to account for formality contrasts
in these pronouns without introducing additional discourse-pragmatic features
into the syntax. Arregi & Nevins (2012) suggest that the feature [formal] could be
used to solve this problem but the fact remains that [formal] is not well-defined
as a feature. That is, it is not readily apparent what is meant by formal. At best the
feature is a useful heuristic device to talk about the facts but does not provide a
principled explanation thereof.

    We account for the distribution of second-person pronouns in several Spanish
varieties and in Catalan by using features that are independently motivated in other
parts of the grammar. Spanish and Catalan second-person pronouns can encode
both personal and spatial deixis. Second-person pronouns in Spanish and Catalan
can contain locative χ feature that positions the addressee close to or far from the
characteristic space of the author.[3] While this proposal is novel, the idea that pro-
nouns can have highly articulated structures is uncontroversial (see for example
Déchaine & Wiltschko 2002; Giusti 2006). The structure of a second-person pro-
noun in Spanish and Catalan is as in (5).

---

3.   Spatial deictics occurring in non-spatial contexts also occur in other languages. In Taba, an
Austronesian language, the acquaintance with topics in the common ground (CG) is a semantic
extension of the meaning *ya* "upward" glossed as REC(OGNITIONAL) (Bowden 2001: 294):

(i)   *odo   lai mo   ne   no-ge   loka   li   ya*
    On the other hand just before PROXIMAL there-ESSIVE banana   LOC REC
    "on the other hand, just before, over there in the bananas… "

Bowden explains that the speaker had already mentioned a banana field earlier allowing it to be
marked with *ya* as it is in the CG.

(5)

```
                NumberP
              /        \
        [±atomic]    ParticipantP
                    /          \
            [+participant]   AuthorP
                            /        \
                        ([χ])      [±author]
```
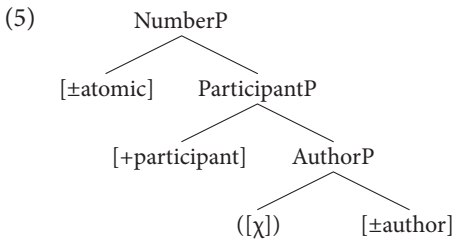
In the structure above, number is either [±atomic] giving plural versus singular contrasts. Second-person always requires [+participant] and χ optionally composes with the individual described by [±author]. The interpretation of [±author] under locative χ falls out from the general meaning of the person features. These features are given in (6) and defined following Halle (1997).

(6) a. [+author] = includes the speaker
     b. [−author] = does *not* include the speaker
     c. [+participant] = includes a(t least one) discourse participant

When χ composes with [+author] it means 'is in the vicinity of the author' and when it composes with [−author] it means the opposite. When formality is marked, χ must appear but it is not employed in every pronoun and thus pronouns vary in whether they include χ at all. For this reason, χ is an optional feature of $Author_0$.

Since we need [±author] in certain representations we must ensure we pick out the addressee and not the author. Following Harbour's (2016) ontology of persons, a discourse has two unique individuals: the author (*i*) and the addressee (*u*) and an arbitrary number of third persons (*o, o', o''*). The combination of [χ, ±author] supplies the psychological location of near (or not near) the author, *i*, and therefore [+participant] can only felicitously pick out the addressee, *u*. Were this not the case, then we would expect the author to be near themselves, which amounts to a tautology (i.e., *i* near *i*), or cases where the author is paradoxically not near themselves (i.e., *i* not near *i*). Given the definition in (6c) a relation of proximity of *u* to *i* is the only well-formed interpretation.

Under this analysis, second-persons that are near the author of the utterance (i.e., [χ, +author]) are interpreted as informal, while second-persons that are not near the author of the utterance (i.e., [χ, −author]) are interpreted as formal. Second-persons not relativized to distance are neutral when there are no contrasting (in)formal forms or they get a default interpretation based on contrasting (in)formal forms. This system allows for a three-way distinction in second persons based on their distance from the author.

6.   Morphology

Having addressed the syntax and semantics/pragmatics of second-person pronouns, we now turn to their morphology. In Distributed Morphology, the narrow syntax does not operate on traditional lexical items, but rather builds structures by combining formal features via the operations Move and Merge (Halle & Marantz 1993; Harley & Noyer 1999). The term *morpheme* refers to a syntactic terminal node and its content (e.g., a bundle of formal features), not to the phonological expression of that terminal node which is known as an exponent. After syntax has built a morphosyntactic representation (MSR), phonological exponents are inserted in a process termed Vocabulary Insertion.

Vocabulary Insertion is subject to underspecification, meaning that Vocabulary Items (VI) need not be fully specified for the morphemes where they can be inserted. Since VIs can be underspecified while morphemes are fully specified, the MSR is independent of the morphophonological representation. Vocabulary Insertion is also subject to the Subset Principle (Halle 1997) which states that a VI can be inserted into a morpheme if it matches all or a subset of the specified grammatical features. If multiple VIs can be inserted into the same morpheme (if they compete) then, the VI that matches the most features of a given morpheme is chosen. However, insertion is blocked if a VI contains features that are not present in the morpheme (see Halle & Marantz 1993; Harley & Noyer 1999).

6.1   A sampling of formality in Spanish and Catalan

6.1.1   *Standard Peninsular Spanish*

SPS is a relatively straight-forward case. Recall that SPS contrasts for formality in both the singular and the plural. Based on our analysis in § 5, we propose the syntactic structures in (7).[4]

(7)   Structural Analysis:
a.   SG/PL. INFOR: $[_{\text{NumP}} \pm\text{atomic} \, [_{\text{participantP}} +\text{participant} \, [_{\text{authorP}} \chi +\text{author}]]]$
b.   SG/PL. FOR: $[_{\text{NumP}} \pm\text{atomic} \, [_{\text{participantP}} +\text{participant} \, [_{\text{authorP}} \chi -\text{author}]]]$

The differences in formality and number arise from the difference in χ's relation to the [author] feature and the difference in the [atomic] feature, respectively. The informal second-person pronouns are [χ, +author] and the formal pronouns are

---

4.   As our structures only represent second-person pronouns, we ignore glossing this throughout. Other glosses used are: FOR(MAL), INFOR(MAL) and FAM(ILIAR). Slashes are used to indicate multiple categories such as SG/PL for 'either' singular or plural.

[χ, −author]. The contrast between [+atomic] and [−atomic] separates the singular pronouns from the plural ones. In (8) we propose several vocabulary items.

(8) Vocabulary Items:
    [χ +author]   ↔   *vosotro* /_ [−atomic]
    [χ +author]   ↔   *tú*
    [χ −author]   ↔   *usted*
    [−atomic]   ↔   *-(e)s*

The VIs for *tú* and *usted* are fairly straightforward. However, there is a distinct informal second-person stem, *vosotro*, that is contextually restricted to [−atomic] contexts. This stem allomorph bocks the otherwise expected **tús* which is ungrammatical. Last, the plural exponent is realized as *-(e)s*; the bracketed *e* surfaces if the stem the plural exponent attaches to ends in a consonant.

6.1.2   *Standard Latin American Spanish*

SLAS primarily differs from SPS because of the formality syncretism in the plural and it lacks the pronoun *vosotros*. These differences notwithstanding, the structures for SLAS in (9) are the same as the ones in (7) for SPS which are repeated here for convenience.

(9) Structural Analysis:
    a.   SG/PL. INFOR: [$_\text{NumP}$ ±atomic [$_\text{participantP}$ +participant [$_\text{authorP}$ χ +author]]]
    b.   SG/PL. FOR: [$_\text{NumP}$ ±atomic [$_\text{participantP}$ +participant [$_\text{authorP}$ χ −author]]]

The noted differences between SLAS and SPS are primarily formalized in the VIs used to spell out these structures. In (10) we propose the vocabulary items for this variety of Spanish.

(10) Vocabulary Items:
    [+atomic, χ, +author]   ↔   *tú/vos*
    [+participant]   ↔   *usted*
    [−atomic]   ↔   *-(e)s*

In (10) the VI for *tú/vos* must be specified with [+atomic] to ensure that it fails to win the competition on vocabulary insertion for an informal plural second-person pronoun structure (i.e., [−atomic, +participant χ +author]). The [+atomic] feature causes the VI for *tú/vos* to be overspecified for this morpheme as the values for atomicity clash. Were this not the case, the VI for *tú/vos* would be inserted as the most highly specified item, producing the ungrammatical **tús* in varieties that use this pronoun. For varieties that predominantly use *vos* the result, **voses* would be equally ungrammatical. In the plural, there is a syncretism and *ustedes* is used for all formality configurations. Our analysis accounts for this syncretism with

underspecification, which Harley (2008) notes is preferable to impoverishment when possible. The VI for *usted* must be underspecified as [+participant] to allow it to be used in all formal (i.e., [χ −author]) contexts both singular and plural and crucially in the informal plural context (i.e., [+atomic, +participant, χ +author]).

6.1.3   *Salvadoran Spanish*

The structural analysis of SS remains largely unchanged from SLAS, the most notable difference is that SS has a three-way formality contrast. The analysis is presented in (11).

(11) Structural Analysis:
  a. SG/PL. FAM: [$_{NumP}$ ±atomic [$_{participantP}$ +participant [$_{authorP}$ χ +author]]]
  b. SG/PL. INFOR: [$_{NumP}$ ±atomic [$_{participantP}$ +participant [$_{authorP}$ −author]]]
  c. SG/PL. FOR: [$_{NumP}$ ±atomic [$_{participantP}$ +participant [$_{authorP}$ χ −author]]]

Here the main difference to SLAS is the addition of an informal second-person pronoun as in (11b) that does not include locative χ at all. Due to this third level of formality, some changes to the VIs we proposed for SLAS are necessary. We analyze the morphological facts for SS with the VIs in (12).

(12) Vocabulary Items:
  [+atomic, χ, +author]   ↔ *vos*
  [+atomic, −author]   ↔ *tú*
  [+participant]   ↔ *usted*
  [−atomic]   ↔ *-(e)s*

Unlike SLAS which uses either *tú* or *vos*, SS uses both of these pronouns and *vos* takes on a different meaning. Here it is familiar whereas in SLAS it is simply informal. This difference aside, SS encounters the same issue of syncretism that was previously noted for SLAS. This is also a case that can be accounted for with our previously proposed underspecification of the VI for *usted*. Consequently, for the same reasons mentioned for SLAS, the VIs for both *tú* and *vos* must be specified with [+atomic] to ensure that they fail to win the competition on insertion for any plural second-person structure (e.g., [−atomic, +participant, χ +author] or [−atomic, +participant, −author]).

The interpretation of the second-person pronouns in SS falls out from a contrastive organization of the relevant features. This notion of contrastive hierarchy in morphosyntax follows from Cowper & Hall (2019) who treat morphosyntactic hierarchies as contrastive hierarchies (e.g., Dresher 2009) instead of feature geometries. They argue that feature dependencies and their crosslinguistic variation are the results of a generic system for acquiring language-specific contrasts. The contrastive hierarchy in (13) illustrates how the various designations of formality obtain in SS.

(13) Contrastive Hierarchy of Formality
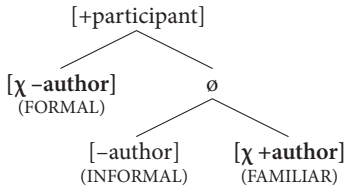
In (13) the notions of formality and familiarity are represented by marked features, with [χ −author], the feature marking formality, taking scope over the non-formal contrast. The other marked feature, [χ +author], is only contrastive among categories lacking formality. [χ +author] gets the interpretation of familiar and the 'informal' designation obtains via contrast. That is, informal arises as an intermediate designation that is neither formal nor familiar. Second persons not relativized to distance are neutral if there are no contrasting (in)formal forms (see § 6.1.5 on Cuban Spanish), otherwise they get a contrastive interpretation, as is the case here.

6.1.4 *Catalan*

Like SS, Catalan pronouns generally exhibit a three-way formality distinction. However, in the plural, only a subset of these contrasts is marked. Our proposal for their structure is given in (14).

(14) Structural Analysis
    a.   SG/PL. INFOR: $[_{\text{NumP}} \pm$atomic $[_{\text{participantP}} +$participant $[_{\text{authorP}} \chi +$author$]]]$
    b.   SG/PL. RESP: $[_{\text{NumP}} \pm$atomic $[_{\text{participantP}} +$participant $[_{\text{authorP}} -$author$]]]$
    c.   SG/PL. FOR: $[_{\text{NumP}} \pm$atomic $[_{\text{participantP}} +$participant $[_{\text{authorP}} \chi -$author$]]]$

Here the main differences to SS are the social meanings that the various structures licence and, of course, the VIs themselves since we are dealing with a different language. We analyze the morphological facts for Catalan in (15).

(15) Vocabulary Items:
    [+atomic, χ, +author]   ↔ *tu*
    [+participant, χ, −author] ↔ *vostè*
    [+atomic, −author]   ↔ *vós*
    [+participant]   ↔ *vosaltre*
    [−atomic]   ↔ *-s*

The first part of the analysis for Catalan should already be quite familiar as we have seen it before in both SLAS and SS. The VI for *vosaltre* must be underspecified as [+participant] to allow it to be used in the two possible non-formal plural contexts (i.e., [−atomic, +participant, χ, +author] and [−atomic, +participant, −author]). This accounts for the observed formality syncretism. The VIs for *tu* and *vós* must

be specified with [+atomic] to ensure that they fail to win the competition on insertion for all plural second-person morphemes. The VI for *vostè* is more specific than the VIs for *vosaltre* or *vós* and therefore *vostè* would win a competition where both items could be inserted into the same morpheme (i.e., [−atomic, +participant, χ, −author] or [+atomic, +participant, χ, −author]) given the subset principle.

Like SS, the interpretation of Catalan's second-person pronouns also falls out from a contrastive organization of the features that make them up. This is illustrated in (16).

(16) Contrastive Hierarchy of Formality

```
            [+participant]
            /           \
    [χ −author]          ø
     (FORMAL)           /  \
               [−author]    [χ +author]
              (RESPECTFUL)   (INFORMAL)
```
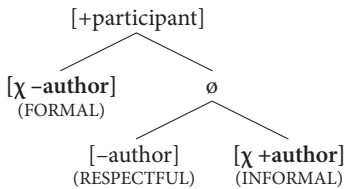
In (16) the notions of formality and informality represent marked features with [χ −author], the feature marking formality, taking scope over the non-formal contrast. Here [χ +author] is only contrastive among categories that lack formality. [χ +author] gets an informal interpretation and the designation of 'respectful' arises through contrast with the informal meaning. This is the same underlying system demonstrated for SS, though the social meanings of the pronouns are different. These levels of formality are modelled on the contrast between different levels of speech (i.e., casual vs. formal speech vs. honorific) in Japanese. Norio Ota (p.c.) points out that formality is a categorization that emerges out of contrast with casual and honorific speech. For formality, indicating neither the marked options of honorific nor casual is more neutral. Essentially, the most formal thing you can do is to neither over nor understate your position in the social hierarchy.

6.1.5 *Cuban Spanish*

Contemporary Cuban Spanish uses the 'informal' pronoun *tú* in all cases, including those where the formal pronoun *usted* would be common in other varieties (Lipski 2007: 258). Therefore, Contemporary Cuban Spanish represents a variety that does not include χ at all. The analysis of second-person pronouns is given in (17).

(17) Structural Analysis:
      SG/PL: $[_{\text{NumP}}$ ±atomic $[_{\text{participantP}}$ +participant $[_{\text{authorP}}$ −author]]]

As we see in (17), in Cuban Spanish the relevant contrast in the structure of its second-person pronouns relies on number (i.e., [±atomic]) and there is no morphologically marked formality in this variety.[5] Accordingly, we propose the vocabulary in (18) to account for the morphology.

(18) Vocabulary Items:
       [−author]   ↔ *usted* /_[−atomic]
       [−author]   ↔ *tú*
       [−atomic]   ↔ *-(e)s*

Here all second-person VIs are specified as [−author]. *Tú* is the more general case and *usted* is a stem allomorph of *tú* that occurs in the context of the number feature [−atomic]. This stem allomorph blocks the otherwise expected **tús* which is ungrammatical. As in all varieties of Spanish plural number is expressed as *-(e)s*. Since there are no contrasting pronouns in this second person system, the sole pronoun available is taken to be neutral rather than informal.

This last case closes out our exploration of second-person pronouns in several varieties of Spanish and Catalan. We have shown that much ground can be covered with a single abstract system that takes formality to be social distance.

7. Conclusion

This chapter has argued for a structure that permits second-person pronouns to be relativized to the characteristic space of the author of speech using the χ operator. This relativization provides the semantic basis for encoding formality as social distance in the morphosyntax and obviates the need for nebulous features such as [FORMAL]. Instead, we take advantage of independently motivated features argued to constitute pronouns crosslinguistically. Our analysis can account for anywhere between zero to three formality contrasts in the singular and in the plural. The approach also accounts for all logical possibilities for encoding formality contrasts in the second person: (i) no formality at all, (ii) formality in the singular but not in the plural paradigm, (iii) formality in both the singular and the plural, (iv) and formality in the plural but not the singular. This last option is conceptually coherent but, to the best of our knowledge, unattested.

---

5. We do not mean to imply that Cuban Spanish has no strategy for expressing formal contrasts but rather the system currently operates like English where this contrast is not marked in the pronoun itself.

References

Arregi, Karlos & Andrew Nevins. 2012. *Morphotactics: Basque auxiliaries and the structure of spellout*. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-3889-8

Bjorkman, Bronwyn, Elizabeth Cowper, Daniel Currie Hall & Andrew Peters. 2019. Person and deixis in Heiltsuk pronouns. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 64(4). 1–18. https://doi.org/10.1017/cnj.2019.13

Bowden, John. 2001. *Taba: Description of a South Halmahera language*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511813085

Butt, John & Carmen Benjamin. 2000. *A new reference grammar of modern Spanish* (3rd edn.). Chicago, IL: NTC Publishing.

Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency. A case study of the three classes of pronouns. In Henk van Riemsdijk (ed.), *Clitics in the languages of Europe*, 145–233. Berlin: Mouton-De Gruyter. https://doi.org/10.1515/9783110804010.145

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris Publications.

Clements, George N. 1985. The geometry of phonological features. In Colin J. Ewen & John M. Anderson (eds.), *Phonology Yearbook* 2. 25–52. Cambridge: Cambridge University Press. https://doi.org/10.1017/S0952675700000440

Cowper, Elizabeth & Daniel Currie Hall. 2019. Scope variation in contrastive hierarchies of morphosyntactic features. In David W. Lightfoot & Jonathan Havenhill (eds.), *Variable properties in language: Their nature and acquisition*, 27–41. Washington, DC: Georgetown University Press. https://doi.org/10.2307/j.ctvfxv99p.7

Déchaine, Rose-Marie & Martina Wiltschko. 2002. Decomposing pronouns. *Linguistic Inquiry* 33. 409–443. https://doi.org/10.1162/002438902760168554

Dresher, B. Elan. 2009. *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511642005

Giusti, Giuliana. 2006. Parallels in clausal and nominal periphery. In Mara Frascarelli (ed.), *Phases of interpretation*, 151–172. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110197723.3.163

Halle, Morris. 1997. Distributed morphology: Impoverishment and fission. In Benjamin Bruening, Yoonjung Kang & Martha McGinnis (eds.), *MIT Working Papers in Linguistics: PF: Papers at the Interface*, 425–449. Cambridge, MA: MITWPL.

Halle, Morris & Alec Marantz. 1993. Distributed Morphology and the pieces of inflection. In Ken Hale & Samuel Jay Keyser (eds.), *The view from Building 20: Essays in linguistics in honour of Sylvain Bromberger*, 111–176. Cambridge, MA: The MIT Press.

Harbour, Daniel. 2011. Descriptive and explanatory markedness. *Morphology* 21. 223–245. https://doi.org/10.1007/s11525-010-9167-0

Harbour, Daniel. 2014. Paucity, abundance, and the theory of number. *Language* 90. 185–229. https://doi.org/10.1353/lan.2014.0003

Harbour, Daniel. 2016. *Impossible persons*. Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/9780262034739.001.0001

Harley, Heidi. 2008. When is a syncretism more than a syncretism? Impoverishment, metasyncretism, and underspecification. In David Adger, Susana Béjar & Daniel Harbour (eds.), *Phi Theory: Phi-features across modules and interfaces*, 251–294. Oxford: Oxford University Press.

Harley, Heidi & Rolf Noyer. 1999. Distributed morphology. *Glot International* 4. 3–9.

Harley, Heidi & Elizabeth Ritter. 2002. Person and number in pronouns: A feature geometric analysis. *Language* 78. 482–526. https://doi.org/10.1353/lan.2002.0158

Lipski, John M. 2007. *El español de America* [American Spanish]. Madrid: Catedra.

Noyer, Rolf. 1992. Features, positions and affixes in autonomous morphological structure. Cambridge, MA: MIT dissertation.

Sells, Peter. 1987. Aspects of logophoricity. *Linguistic Inquiry* 18. 445–479.

Wheeler, Max, Nicolau Dols, & Alan Yates. 1999. *Catalan: A comprehensive grammar*. London: Routledge. https://doi.org/10.4324/9780203300275

SECTION D

Bridging issues in linguistics

Cyclical change in affixal negation

Elisabeth Gibert-Sotelo

Universitat Rovira i Virgili

This chapter offers a contrastive analysis of the negative prefix *iN-* in Latin and Romance that shows that this element has undergone a reanalysis throughout the evolution. In particular, it is proposed that *iN-* has evolved from an adjunct to a categorizing affix, a change that partially fits the negative cycle 'adjunct > specifier > head > affix' (van Gelderen 2011). The proposed analysis builds, on the one hand, on Newell's (2008) morphophonological approach to English negative prefixes; and, on the other hand, on De Clercq's (2013) and De Clercq & Vanden Wyngaerd's (2017) work on affixal negation. Alongside, the paper endeavors to show the advantages of using the Nanosyntax model when dealing with diachronic variation at the syntax-lexicon interface.

Keywords: affixal negation, prefix *iN-*, negative cycle, Nanosyntax, Latin, Romance languages

1.   Introduction

In this study I argue that (partial) cyclical change can be traced in affixal negation and,[1] specifically, in the evolution of the negative prefix *iN-* from Latin to Romance languages.[2] At first glance, this prefix seems to behave alike in both linguistic systems: its negative meaning has been kept unaltered and it is basically attested in adjectival predicates (Brea 1976; Montero Curiel 1999; Gibert-Sotelo 2017). However, *iN-* shows a higher degree of productivity in Latin than in Romance. Consider, for instance, the ability of this prefix to combine with all sorts of adjectival

---

[1]   The label 'affixal negation' (Zimmer 1964) refers to the use of derivative affixes to encode negation.

[2]   The notation *iN-* is here used to encompass the allomorphic variants of the prefix (Baldi 1989; Horn [1989] 2001; De Clercq 2013; among others). This notation has also been employed to distinguish this negative prefix from the homonymous Latin prepositional prefix *in-* "in" (Gibert-Sotelo 2017: 195).

passive participles in Latin vs. the impossibility of doing so in a Romance language like Spanish:[3]

(1)   a. *in-amatus* "unloved" (Latin)   vs.   **in-amado* (Spanish)
      b. *in-castigatus* "unpunished" (Latin)   vs.   **in-castigado* (Spanish)
      c. *in-cenatus* "dinnerless" (Latin)   vs.   **in-cenado* (Spanish)
      d. *in-dictus* "unsaid" (Latin)   vs.   **in-dicho* (Spanish)
      e. *in-/ir-ruptus* "unbroken" (Latin)   vs.   **in-/ir-roto* (Spanish)

[Based on Gibert-Sotelo 2017: 332 (75)]

Inspired by insights in Newell (2008), I propose that the evolution from Latin to Romance entailed the reanalysis of the negative prefix *iN-*, which evolved from a morphological adjunct to a categorizing affix. In a nanosyntactic approach to grammar as pursued in this paper, morphemes can spell out multiple terminal nodes (Phrasal Spell-Out). Accordingly, and taking into account De Clercq's (2013) and De Clercq & Vanden Wyngaerd's (2017) decomposition of affixal negative markers, I assume that *iN-* lexicalizes a Neg(ation) feature as well as a Q(uantifier) feature both in Latin and Romance, which accounts for its negative meaning and for its requirement to only combine with gradable predicates. The difference between Latin and Romance, therefore, is that Romance *iN-* also lexicalizes a categorizing *a*(djectival) feature that forces it to be merged as a sequence of heads on top of the predicate it negates, categorizing this predicate as an adjective and hence preventing the addition of the prefix to nouns or verbs. Since Latin *iN-* does not contain a categorizing *a* feature, it can be merged at an adjunct position, thus showing fewer restrictions in its combinatorial patterns.

The change undergone by the negative prefix *iN-* in its evolution from Latin to Romance partially fits the negative cycle as described by van Gelderen (2011: 17 (23)) and represented in (2), which is further evidence of the adequacy of approaching the language faculty in terms of cyclical change (van Gelderen 2011: 3).

(2)   Adjunct > Specifier > Head > Affix

The paper is organized in five sections, including this introduction. In § 2 I show the different behavior of the prefix *iN-* in Latin and Romance. Section 3 summarizes the possibilities offered by a theoretical framework such as Nanosyntax in the account of diachronic change at the syntax-lexicon interface. In § 4 a nanosyntactic analysis of *iN-* in Latin and Romance is offered which naturally derives its different behavior and agrees with van Gelderen's (2011) approach to the negative cycle. Finally, § 5 concludes the paper.

---

3.   I mainly focus on Catalan and Spanish data, but include references to other Romance languages as necessary.

2. Latin vs. Romance *iN-*

As already pointed out, the degree of productivity of *iN-* is higher in Latin than in Romance. This prefix is strictly restricted to adjectival predicates in Romance languages, and its addition to nouns and verbs systematically leads to ungrammaticality.[4] In fact, *iN-* prefixation is commonly used as a test to distinguish verbal participles from adjectival ones, since verbal participles systematically reject *iN-*prefixation but adjectival participles may allow it (Varela 1983, 1990, 2002; Bosque 1990, 1999; Oltra-Massuet 2014):

(3) Verbal participles disallow *iN-* prefixation
  a. *L'obra   va   ser   (*in)acabada   pel   seu deixeble.*
     the=work PRF.3SG be.INF IN-finish.PTCP.PST by=the his disciple
     "The work was (*un)finished by his disciple."   (Catalan)
  b. *El río   ha sido (*in)contaminado   por los vertidos.*   (Spanish)
     the river has been IN-pollute.PTCP.PST by the spillage.PL
     "The river has been (*un)polluted by the spillage."

(4) Adjectival participles allow *iN-* prefixation
  a. *L'obra   està (in)acabada.*   (Catalan)
     the=work is   IN-finish.PTCP.PST
     "The work is (un)finished."
  b. *El río   está (in)contaminado.*   (Spanish)
     the river is   IN-pollute.PTCP.PST
     "The river is (un)polluted."

By contrast, Pinkster (2015: 734) points out that in Early Latin *iN-* can be added to present participles that seem to keep their verbal character, thus showing a less restrictive behavior than its Romance descendant:[5]

---

4. Some apparent counterexamples are attested. Consider the Catalan noun *imparcialitat* "impartiality" or the Spanish verb *inactivar* "to inactivate", both embedding the negative prefix. Crucially, these words do not involve the addition of *iN-* to a nominal or verbal base, respectively, but the addition of the prefix to an adjectival base. *Imparcialitat* "impartiality" is a deadjectival noun derived from the *iN-* prefixed adjective *imparcial* "impartial"; and the verb *inactivar* is not the result of combining the negative prefix with the verb *activar* "to activate", but it is a deadjectival verb created upon the *iN-* prefixed adjective *inactivo* "inactive", given that its meaning, rather than "to not activate", is "to make become inactive".

5. Pinkster (2015: 734) mentions that affixation of *iN-* to verbal participles goes out of use in Classical Latin, except for certain expressions that are used as set phrases, like *me insciente* "without my knowing it". Along these lines, an anonymous reviewer observes that *indicente me* (cf. (5)) is listed as a set phrase in Bornecque & Cauët's (1941) dictionary. In fact, set phrases reflect phenomena that were productive in previous stages of the language, which suggests that *iN-* could combine with verbal participles in the first stages of Early Latin.

(5)   *Non me   in-dicente   haec   fiunt.*   (Latin)
      not I.ABL IN-say.PTCP.PRS.ABL this.NOM.PL be_made.PRS.3PL
      "I did not fail to predict that this would happen."
                          [Ter. *Ad.* 507; *apud* Pinkster 2015: § 8.51, 734 (a)]

That *iN-* prefixation is not strictly restricted to adjectival predicates in Latin is also evidenced by the fact that some *iN-* prefixed verbs are documented in Latin Dictionaries (although they are very few and dictionaries tag them as rare or infrequent). An example of this is the verb *indecere* "to misbecome", "to not be suitable", derived from *decere* "to be suitable" (cf. Lewis & Short 1879 and Gaffiot 1934, *s.v.*).[6]

Another crucial distinction found between Latin and Romance *iN-* is that the former shows a more autonomous behavior. Hence, as documented by Pinkster (2015: 735), in Latin poetry the negative prefix *iN-* can be separated from the predicate it negates for metrical reasons (a rhetorical phenomenon known as *tmesis*). This is exemplified in (6), where the coordinator *-que* is added between the prefix *iN-* and *merentes*, the negated predicate:

(6)   *… saepe nocentes /   praeterit   exanimat=que*
      often   hurt.PTCP.PRS.ACC.PL pass_by.PRS.3SG kill.PRS.3SG=and
      in-dignos   in=que merentes?   (Latin)
      IN-worthy.ACC.PL IN=and deserve.PTCP.PRS.ACC.PL
      "… often passes the guilty by and slays the innocent and undeserving?"
                          [Lucr. 2.1103-4; *apud* Pinkster 2015: § 8.51, 735 (c)]

Further evidence of the higher degree of autonomy of this prefix in Latin is found in its phonological behavior. Baldi (1989: 6) notices that assimilation of the nasal consonant is optional in Latin: the unassimilated form [in] is available in every environment (7). Conversely, *iN-* always assimilates the nasal to the following consonant in Romance languages (8):[7]

---

6. Notice that, unlike the Romance apparent counterexamples, like Spanish *inactivar* "to make become inactive" (cf. footnote 4), the prefix seems to take scope over the verbal predicate in the Latin verbs. Hence, Latin *indecere* can be paraphrased as "to not be suitable", contrary to Spanish *inactivar*, which cannot be paraphrased as "to not activate".

7. The assimilation rule is not always productive in current French. In particular, new *iN-* prefixed adjectives involving *-ble* suffixation usually show no assimilation: *inlassable* "tireless", *inracontable* "impossible to tell". The phonological behavior of this prefix in French has been studied by Zimmer (1964); Tranel (1976); Apothéloz (2003); and Buchi (2012), among others. I thank an anonymous reviewer for bringing this point to my attention.

(7) Latin
inermis "unarmed", *inmodestus* "unrestrained", *inlacerabilis* "that cannot be torn", *inreverens* "irreverent"   [Baldi 1989: 6]

(8) a. Spanish: *irreverente* "irreverent", *imperfecto* "imperfect", *ilegal* "illegal"
    b. Catalan: *irreverent* "irreverent", *imperfecte* "imperfect", *il·legal* "illegal"
    c. French: *irrévérent* "irreverent", *imparfait* "imperfect", *illégal* "illegal"
    d. Italian: *irriverente* "irreverent", *imperfetto* "imperfect", *illegale* "illegal"
    e. Portuguese: *irreverente* "irreverent", *imperfeito* "imperfect", *ilegal* "illegal"

To sum up, Latin *iN-* is more productive and autonomous than Romance *iN-*, showing less restrictions on the bases it attaches to, allowing material to be inserted between the prefix and the lexical base, and not systematically assimilating the nasal consonant.

3.   A nanosyntactic approach to the syntax-lexicon interface

Based on a recent neo-constructionist theory known as Nanosyntax (Starke 2009; Caha 2009; Fábregas 2007, 2016; Pantcheva 2011; Baunaz et al. 2018), I adopt the view that words are syntactically built. Like other syntactically-oriented models, Nanosyntax proposes an architecture of grammar in which the lexicon is accessed after syntax (late insertion). The lexicon, therefore, is not generative in any sense, and its unique function is to provide lexical exponents in order to spell out the syntactic structure. What distinguishes Nanosyntax from other neo-constructionist models (e.g., Distributed Morphology) is the assumption that the relation between syntax and the lexicon is direct, which leaves no place for an intermediate morphological level independent from syntax: syntax generates configurations that can be spelled out as morphemes, words or phrases (Starke 2009: 6).

A straightforward consequence of this approach to the syntax-lexicon interface is Phrasal Spell-Out, that is, the possibility of inserting lexical exponents into phrasal nodes that span multiple terminals, as represented in (9). For a lexical exponent to match a syntactic node, its lexical entry must contain a superset of the features contained in the syntactic node (Superset Principle). This means that lexical exponents can spell out the whole set of features they are specified for (9a) or only a subpart thereof (9b). The Superset Principle is constrained by conditions on 'underassociation', as proposed by Ramchand (2008: 97–99). Hence, a lexical item can leave an unmatched feature only if this feature is independently identified by another lexical item within the same domain, as is the case in (9b).

(9) a. *To the shop*
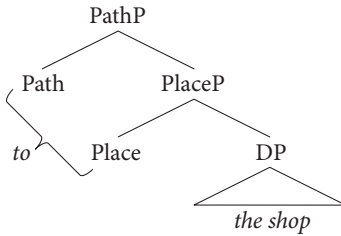
  b. *Into the shop*
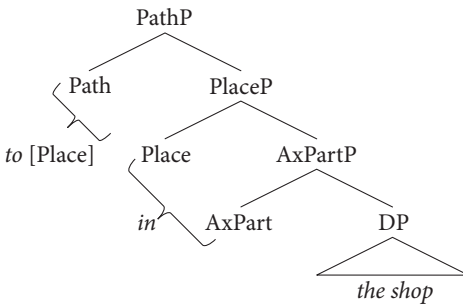
[Adapted from Gibert-Sotelo 2017: 19 (6) and 22 (11)]

As conforming to Phrasal Spell-Out, the preposition *to*, which expresses direction "toward" as well as final location "at", can spell out all the features it is specified for, that is, Path and Place, as represented in (9a). By the Superset Principle, though, *to* can underassociate and spell out only a subpart of its features. This is what occurs in (9b), where *to* underassociates its Place feature because it combines with a purely locative preposition, *in*, that identifies Place – in addition to AxPart, a feature that specifies that the location is inside the reference object (see Svenonius 2010 for the structure of PPs). Following Ramchand's notation, the Place feature that *to* leaves underassociated is shown in brackets.

A major achievement of the Nanosyntax framework is that it provides tools to account for language variation: if language variation corresponds to different ways of lexicalizing the syntactic structure, then it can be defined as size differences among the trees spelled out by lexical exponents in various languages (Starke 2014). Such a view not only accounts for synchronic crosslinguistic diversity, but also for diachronic change, given that the amount of structure lexicalized by a lexical item may vary through the history of the language (a nanosyntactic idea that has been adopted by Mendívil-Giró 2015). Hence, and as pointed out in Gibert-Sotelo (2017), diachronic morphosyntactic change occurs when the relation between one underlying structure and one lexical exponent is reinterpreted. Reanalysis can

imply (1) that a morpheme adds a feature to its spell-out, or (2) that it loses one of its features. As will be shown below, the former case explains the different behavior of the negative prefix *iN-* in Latin and Romance.[8]

4. Proposal

The divergences existing between Latin and Romance *iN-* are in many respects parallel to those observed by Newell (2008) between *un-* and *iN-* in English. In particular, she deals with the following phonological and syntactic contrasts, which are taken as evidence of the different status of these two negative prefixes. First of all, the nasal of *iN-*, unlike that of *un-*, systematically assimilates to the following consonant, as shown in (10). Newell (2008: 181) argues that this difference is due to the fact that *iN-* is spelled out in the same phase domain as its sister – and, therefore, in the same phonological domain –, whereas *un-* is not.[9]

(10)   *impolite*   vs. *unpopular*
       illegal   vs. *unlucky*
       irrational vs. *unreal*

---

8.   A case of grammatical change that involves losing features is found in Gibert-Sotelo (2017, 2021) with regard to Latin and Romance directional prefixes. Specifically, it is argued that the Latin prefixes *ex-*, *de-* and *dis-* lost the lowest projections of their structure along their evolution to Romance *(d)es-*. Interestingly, the features lost by the directional prefixes were the more 'lexical' ones: AxPart, which provides information regarding the topological properties of the initial or final location, and the root ($\sqrt{}$) node, which is the locus of conceptual content (Acedo-Matellán & Mateu 2014). By contrast, and as will be made clear in the following, the node added to the structure of *iN-* is a grammatical one: a category-assigning head (notice that I adhere to the view, adopted in Distributed Morphology, that categorizing heads are functional; cf. Marantz 1997, 2001). We can therefore reformulate our claim and hypothesize that grammaticalization implies (1) that a lexical exponent adds a functional (grammatical) feature to its lexically stored tree, or (2) that it loses a lexical feature and keeps the more grammatical one(s) (see Mendívil-Giró 2016 for a similar view).
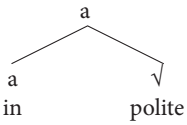
9.   Within level-ordered morphology, Kiparsky (1982, 1983) takes this phonological property (i.e., the (non)assimilation of the nasal) as evidence that *iN-* is to be analyzed as a Level I affix and *un-* as a Level II one. See Siegel (1974); Allen (1978); Selkirk (1982); and Horn ([1989] 2001) for related discussion.

Secondly, *iN-* is restricted to adjectival bases, but *un-* can be added to adjectives (*unhappy*) and also to nouns (*unperson*) and proper nouns (*unBritney*).[10] Thirdly, *iN-*, but not *un-*, may be added to bound roots, and the result is always an adjective (cf. *inept, inane*). Finally, only *un-* can give rise to bracketing paradoxes. An example of this is *unhappier*: its phonological form points to the bracketing in (11a), with the prefix scoping over the comparative suffix – given that the allomorphic variant *-er* would not emerge in a three-syllable word –; on the contrary, its meaning – which is "more unhappy" and not "not more happy" – points to a bracketing where the comparative suffix scopes over the negative prefix, as the one in (11b):
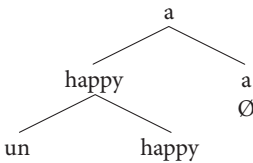
(11)   *unhappier*
     a.   Phonological bracketing: [un[happier]]
     b.   Semantic bracketing: [[unhappy]er]

Newell (2008) accounts for all these differences by assuming that *iN-* is a category-defining head that projects an adjectival label and is therefore merged cyclically (12a), whereas *un-* is a morphological adjunct able to be counter-cyclically inserted (12b):

(12)   a.

     b.

[Newell 2008: 182 (17)]

In the following, I argue that a similar account can be given for Latin and Romance *iN-*, the former being a case of morphological adjunct and the latter a categorizing prefix.[11]

---

10. Newell states that *un-* is also added to verbs, as in *untie* or *undo*. However, the prefix added to verbs in order to encode reversative meaning is different from the negative prefix (Jespersen 1917). Reversative *un-* is the evolution of Old English *on(d)-* (of common descent with Dutch *ont-* and German *ent-*), whereas negative *un-* descends from Old English *un-* (cognate with Latin *iN-*, Greek *a(n)-*, Dutch *on-* and German *un-*). Both prefixes, however, have converged orthographically and phonologically, and they are felt to be semantically connected (Marchand 1969), which has led some authors to offer a unitary analysis of both forms (Maynor 1979; Andrews 1986; Newell 2008). See Horn (2002: 13) for discussion.

11. As noticed by an anonymous reviewer, the analysis of *un-* put forward by Newell (2008) (see (12b) above) predicts that this prefix, being an acategorial adjunct, can be added to any type of word. However, words like **untable* are not attested, which points to the need of establishing

4.1 Nanosyntax of Romance *iN-*

Romance *iN-* shows the same phonological and syntactic restrictions as its English counterpart: it assimilates the nasal to the following consonant (cf. § 2, examples in (8)); it can only be attached to adjectival bases (cf. § 2, examples in (3) and (4)) and, when added to acategorial roots, the result is systematically an adjective (cf. Cat. *incolor* "colorless" or Sp./Cat./Fr./It./Por. *imberbe* "beardless"); and it does not induce bracketing paradoxes of any sort. Accordingly, I assume, in line with Newell (2008), that Romance *iN-* is an adjectivizing affix that lexicalizes a category feature *a*(djective), which explains the adjectival nature of all the words prefixed with *iN-* in Romance languages.

This is not, however, the only feature for which *iN-* is specified. Drawing on the nanosyntactic analysis of negative affixes developed by De Clercq (2013) and De Clercq & Vanden Wyngaerd (2017), I will further assume that *iN-* also lexicalizes a Q(uantifier) feature (cf. Corver 1997) that accounts for its restriction to only combine with gradable bases, and a Neg(ation) feature that is responsible for its negative meaning. Evidence of the presence of a Q feature in the internal structure of *iN-* is provided by the fact that *iN-* prefixation is only available with gradable adjectives (13), its presence in relational adjectives – which are non-gradable – being systematically rejected (14) (cf. Varela & Martín García 1999; Horn [1989] 2001; Costa 2008; among others):

(13) a. *Un libro absolutamente in-traducible*　　　　　　　　(Spanish)
　　　　a　book absolutely　　IN-translatable
　　　　"An absolutely untranslatable book"

　　b. *Una dona　molt in-feliç*　　　　　　　　　　　　　(Catalan)
　　　　a　woman very　IN-happy
　　　　"A very unhappy woman"

　　c. *Um terreno bastante in-explorado*　　　　　　　　　(Portuguese)
　　　　a　terrain pretty　IN-explored
　　　　"A pretty unexplored terrain"

　　d. *Un garçon complètement im-berb-e*　　　　　　　　　(French)
　　　　a　boy　completely　IN-beard-E
　　　　"A completely beardless boy"

(14) a. *Un estado (*in)mental*　　　　　　　　　　　　　　(Spanish)
　　　　a　state　IN-mental
　　　　"A(n) (*un)mental state"

─────────────────

some kind of restriction to the adjunction of *un-*. Even though the present chapter analyzes Latin *iN-* as an adjunct, the account proposed here predicts that it can only combine with bases that admit degree quantification, i.e., with scalar predicates (cf. § 4.2).

b. *L'energia*   (*in)*solar*   (Catalan)
   the=power IN-solar
   "The (*un)solar power"

c. *Un examen*   (*im)*médical*   (French)
   a   check-up IN-medical
   "A(n) (*un)medical check-up"

Therefore, the complete structure that I propose for this prefix in Romance is the one depicted below, which specifies that it is a negative marker (Neg) that quantifies over a scale (Q) and gives rise to adjectival predicates (*a*):

(15)   Romance *iN-*

```
              NegP
           /        \
        Neg          QP
                   /     \
                  Q       aP
                          |
                          a
```
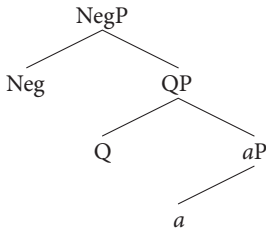
By the Superset Principle, *iN-* can spell out all the features it is specified for, or it can underassociate and spell out a subset of these features. Crucially, the addition of this prefix to adjectival bases systematically involves underassociation of *a*, since this feature is already lexicalized by the adjectival base. In these cases, thus, *iN-* only spells out Neg and Q, whereas *a* is spelled out by the adjective to which *iN-* is attached:

(16)   Combination of Romance *iN-* with adjectival bases (e.g., Spanish *infeliz* "unhappy")

```
                 NegP
              /        \
           Neg          QP
            |         /     \
        iN- [a]      Q       aP
                           /     \
                          a      √FELIZ
                          _____/
                            feliz
```
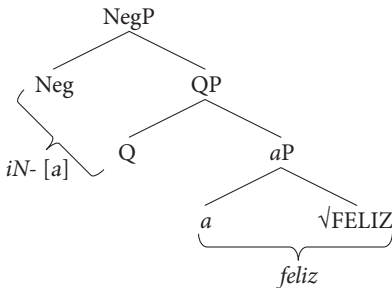
By contrast, when *iN-* is added to acategorial roots (as in Catalan *incolor* "colorless"), it also spells out the adjectivizing *a* feature and categorizes the root as an adjective:

(17) Combination of Romance *iN-* with acategorial roots (e.g., Catalan *incolor* "colorless")

```
                NegP
              /      \
           Neg        QP
            |        /    \
          iN-      Q        aP
                   |      /    \
                  iN-    a    √COLOR
                               ‿‿‿
                               color
```
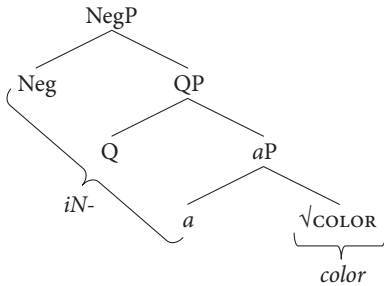
The proposed analysis accounts for the behavior of *iN-* in Romance. First, it cannot combine with nouns and verbs because it bears a categorizing *a* feature that is compatible with adjectival bases or acategorial roots but not with nominal or verbal bases. Second, it assimilates the nasal consonant because it belongs to the same spell-out domain as the negated base. Finally, the impossibility of inserting lexical material between *iN-* and the base follows from the inherent structure of the prefix: if *iN-* lexicalizes a category-defining head, then it must necessarily be inserted cyclically as a sequence of heads on top of the configuration it negates, which precludes any additional head to intervene between both constituents.

4.2   Nanosyntax of Latin *iN-*

Latin *iN-* shows a behavior parallel to that of English *un-*: it must not necessarily assimilate the nasal consonant (cf. § 2, examples in (7)); it is not strictly restricted to adjectival predicates, since it is attested in a few predicates showing verbal character (cf. § 2, Example (5)); and, although bracketing paradoxes have not been documented with this prefix, we have seen that it may license lexical material to be inserted between it and the negated predicate in certain contexts (cf. the case of tmesis in § 2, Example (6)).

The divergences found between Romance and Latin *iN-* are due to the fact that the latter does not lexicalize the *a*(djective) feature, but only the Q(uantifier) and Neg(ation) features:

(18) Latin *iN-*

```
               NegP
             /      \
          Neg        QP
                    /   \
                   Q
```
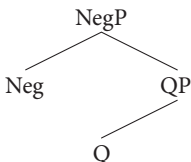
Because of the absence of the categorizing *a* feature in the tree lexicalized by Latin *iN-*, it can be merged in the structure counter-cyclically, as an adjunct. This is illustrated in (19):

(19)   *iN-* prefixed adjectives in Latin
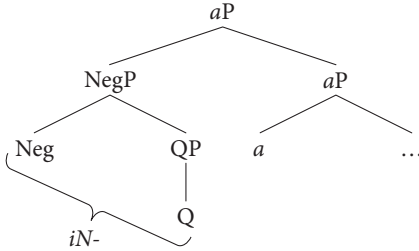
If *iN-* is not merged as a sequence of heads on top of the negated adjective, then the possibility of adding material in between is not banned, which accounts for cases of tmesis like the one pointed out above. Besides, if *iN-* does not contain the categorizing feature, it must not necessarily be involved in adjectival configurations, which would allow its use in environments showing verbal, rather than adjectival, behavior. Finally, if *iN-* is inserted at an adjunct position, then it does not belong to the same spell-out domain as the negated predicate, which explains the possible lack of assimilation of the nasal consonant.

It should be noted, though, that Latin *iN-* cannot be affixed to all kinds of bases. For instance, it is not attested with nouns like *fenestra* "window" or *lapis* "stone". This restriction is predicted in our analysis, given that this prefix lexicalizes a Q feature that forces it to combine with a scalar base over which to quantify. Accordingly, the resulting *iN-* prefixed word must necessarily be a scalar predicate. Hence, even though Latin *iN-* is inserted at an adjunct position and shows more freedom in its combinatorial patterns than Romance *iN-*, it also shows certain restrictions due to the Q feature it is specified for.

Finally, an account should be given for cases in which *iN-* combines with acategorial roots in Latin, since the result of adding this prefix to an acategorial root is always an adjective (cf. *imberbis* "beardless", *implumis* "featherless", *inanimis* "breathless, lifeless"). This could lead to the conclusion that the prefix also includes a categorizing *a* feature in this language. However, *iN-* has been shown to behave as an adjunct in Latin, and this means that it can be adjoined at different positions within the configuration. The most frequently attested cases are those in which *iN-* is added to an adjective (given that predicative adjectives are scalar in nature), but this prefix can also be adjoined low in the structure, at the level of the root, which is in fact the case in *imberbis* "beardless", *implumis* "featherless" or *inanimis* "lifeless" (cf. the non-existence of **berbis* "with a beard", **plumis* "with feathers" or **animis* "full of life").

(20)   Combination of Latin *iN-* with acategorial roots (e.g., *implumis* "featherless")
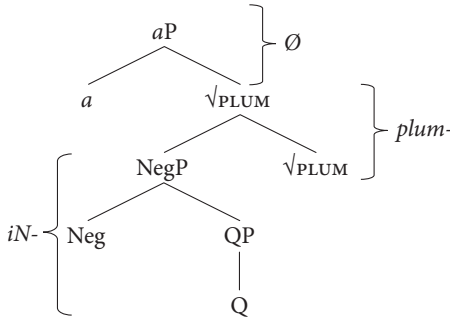
The above-mentioned restriction – namely, that *iN-* can only combine with scalar bases because it contains a Q feature – provides us with an account for these cases. Crucially, it is not any kind of root that allows adjunction of the negative prefix in Latin, but only roots that denote (in)alienable possessions or inherent properties, that is, elements that are scalar and that establish a relation of predication with another entity. As pointed out by Brea (1976: 323), when *iN-* is used to negate a bare root, what is negated is not the existence of the entity denoted by the root, but the relation that this entity establishes with another entity/object (cf. the adjective *implumis*, which does not predicate the non-existence of *plum-* "feathers", but the non-existence of *plum-* "feathers" in a particular being). Therefore, the output of combining *iN-* with a bare root must necessarily be a scalar predicative item, which explains the systematic interpretation of these constructions as predicative adjectives (a fact that could be at the base of the reanalysis of this prefix as an adjectival categorizer).

4.3   From Latin to Romance *iN-*

The evolution from Latin to Romance involved the reanalysis of the negative prefix *iN-*, which evolved from a somehow free element (a morphological adjunct) to a derivative morpheme (a categorizing affix). This reanalysis is illustrated in (21), where XP indicates the possibility of using *iN-* with non-adjectival predicates in Latin:

(21)   Reanalysis of the negative prefix *iN-*

a. Latin   b. Romance
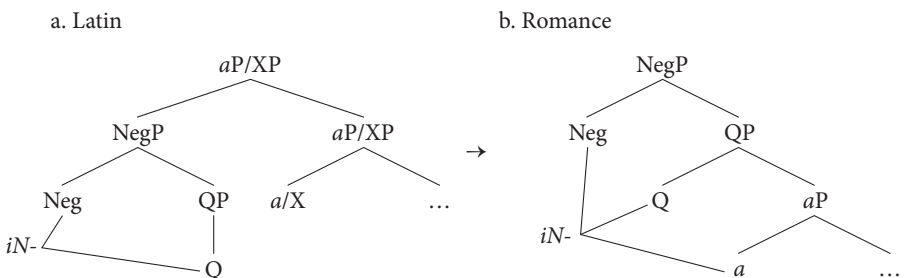
As argued by van Gelderen (2011: 5), grammaticalization (or reanalysis) is "a process whereby lexical items lose phonological weight and semantic specificity and gain grammatical functions". This is, in fact, the process undergone by *iN-* in the evolution from Latin to Romance: it gained a grammatical function (the adjectivizing feature) and lost phonological weight (systematically assimilating the nasal consonant). This change partially reflects the negative cycle as represented in van Gelderen (2011: 17 (23)), which entails reanalysis from adjunct to affix.

(22)   Adjunct > Specifier > Head > Affix

The account offered in (21), in which an adjunct element has become incorporated in the main tree, is a case of reanalysis of pair-merge as set-merge (Chomsky 2000), and this is in accordance with van Gelderen's (2019: 30) observation that the latter is to be preferred over the former.

5.   Concluding remarks

In these pages I have offered a contrastive analysis of the negative prefix *iN-* in Latin and Romance. By doing so, it has been evidenced that this element has altered its grammatical behavior throughout the evolution: it has changed from an adjunct (Latin) to a categorizing affix (Romance), a reanalysis that partially fits the negative cycle 'adjunct > specifier > head > affix' (van Gelderen 2011).

Based on Newell's (2008), De Clercq's (2013) and De Clercq & Vanden Wyngaerd's (2017) approaches to negative affixes, I have decomposed Romance *iN-* into a sequence of three heads: a Neg(ation) head that is the reason for its behavior as a negative marker, a Q(uantifier) head that accounts for its need to exclusively combine with gradable bases, and an *a*(djective) head that forces it to be merged cyclically and explains the systematic adjectival categorization of the predicates negated with *iN-* in Romance languages.

As for Latin *iN-*, it has been argued that it does not involve the same structure as its Romance descendant: it does not lexicalize an adjectival categorizing head and, accordingly, it is not restricted to adjectival predicates, which explains its possible attestation in predicates other than adjectives. The lack of the categorizing head within the internal structure of Latin *iN-* allows this prefix to be counter-cyclically merged as an adjunct, which accounts for its higher degree of morphosyntactic and phonological autonomy.

Finally, Nanosyntax has provided us with useful tools to achieve an accurate analysis of the diachronic variation experienced by *iN-*. Based on the Phrasal Spell-Out principle, I have put forward that in the step from Latin to Romance this prefix changed the size of its lexically stored tree by gaining a grammatical feature: the adjectivizing head.

Acknowledgements

Funding

References

Acedo-Matellán, Víctor & Jaume Mateu. 2014. From syntax to roots: A syntactic approach to root interpretation. In Artemis Alexiadou, Hagit Borer & Florian Schäfer (eds.), *The syntax of roots and the roots of syntax*, 14–32. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199665266.003.0002

Allen, Margaret. 1978. Morphological Investigations. Storrs, CT: University of Connecticut dissertation.

Andrews, Edna. 1986. A synchronic semantic analysis of *de-* and *un-* in American English. *American Speech* 61(3). 221–232. https://doi.org/10.2307/454665

Apothéloz, Denis. 2003. Le rôle de l'iconicité constructionnelle dans le fonctionnement du préfixe négatif *in-*. *Cahiers de linguistique analogique* 1. 35–63.

Baldi, Philip. 1989. Negative prefixation in Latin. *L'information grammaticale* 43. 3–6. https://doi.org/10.3406/igram.1989.1971

Baunaz, Lena, Karen De Clercq, Liliane Haegeman & Eric Lander (eds.). 2018. *Exploring Nanosyntax*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190876746.001.0001

Bornecque, Henri & Fernand Cauët. 1941. *Le Dictionnaire Latin-Français du Baccalaureat*. Paris: Libraire Classique Eugène Belin.

Bosque, Ignacio. 1990. Sobre el aspecto en los adjetivos y en los participios. In Ignacio Bosque (ed.), *Tiempo y aspecto en español*, 177–214. Madrid: Cátedra.

Bosque, Ignacio. 1999. El sintagma adjetival. Modificadores y complementos del adjetivo. Adjetivo y participio. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, Volume 1, 217–310. Madrid: Espasa-Calpe.

Brea, Mercedes. 1976. Prefijos formadores de antónimos negativos en español medieval. *Verba* 3. 319–341.

Buchi, Éva. 2012. *Réel, irréel, inréel*: depuis quand le français connaït-il deux préfixes négatifs *in-*? In Mario Barra-Jover, Guylaine Brun-Trigaud, Jean-Philipe Dalbera, Patrick Sauzet & Tobias Scheer (eds.), *Études de Linguistique gallo-romane*, 323–340. Saint-Denis, France: Presses universitaires de Vincennes.

Caha, Pavel. 2009. The Nanosyntax of case. Tromsø, Norway: University of Tromsø dissertation. http://hdl.handle.net/10037/2203

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on Minimalist Syntax in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Corver, Norbert. 1997. *Much*-support as a last resort. *Linguistic Inquiry* 28(1). 119–164. http://www.jstor.org/stable/4178967

Costa, Sylvia. 2008. *Negación y prefijación negativa. Algunos aspectos de la forma y la interpretación de las secuencias negativas*. Montevideo, Uruguay: Universidad de la República.

De Clercq, Karen. 2013. A unified syntax of negation. Ghent, Belgium: University of Ghent dissertation. http://hdl.handle.net/1854/LU-4215769

De Clercq, Karen & Guido Vanden Wyngaerd. 2017. Why affixal negation is syntactic. In Aaron Kaplan, Abby Kaplan, Miranda McCarvel & Edward J. Rubin (eds.), *Proceedings of WCCFL 34*, 151–158. Sommerville, MA: Cascadilla Press.

Fábregas, Antonio. 2007. The Exhaustive Lexicalisation Principle. *Nordlyd* 34(2). 165–199. https://doi.org/10.7557/12.110

Fábregas, Antonio. 2016. *Las nominalizaciones*. Madrid: Visor Libros.

Gaffiot, Félix. 1934. *Dictionnaire Latin-Français*. Paris: Hachette.

Gelderen, Elly van. 2011. *The linguistic cycle: Language change and the linguistic faculty*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199756056.001.0001

Gelderen, Elly van. 2019. Cyclical change and problems of projection. In Anne Breitbarth, Miriam Bouzouita, Lieven Danckaert & Elisabeth Witzenhausen (eds.), *Cycles in language change*, 13–32. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198824961.003.0002

Gibert-Sotelo, Elisabeth. 2017. Source and negative prefixes: On the syntax-lexicon interface and the encoding of spatial relations. Girona, Spain: Universitat de Girona dissertation. http://hdl.handle.net/10803/461414

Gibert-Sotelo, Elisabeth. 2021. On the encoding of negation by Source prefixes and the satellite-/verb-framed distinction: Evidence from Latin and Spanish. *Linguistic Variation* 21(1). 214–246. https://doi.org/10.1075/lv.00037.gib

Horn, Laurence R. [1989] 2001. *A Natural history of negation*. Stanford: CSLI Publications.

Horn, Laurence R. 2002. Uncovering the un-word: A study in lexical pragmatics. *Sophia Linguistica* 49. 1–64.

Jespersen, Otto. 1917. *Negation in English and other languages*. Copenhagen: A.F. Høst.

Kiparsky, Paul. 1982. Lexical Morphology and Phonology. In The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, 3–91. Seoul: Hanshin.

Kiparsky, Paul. 1983. Word Formation and the Lexicon. In Frances Ingemann (ed.), *Proceedings of the 1982 Mid -America Linguistics Conference*, 3–29. Lawrence, KS: University of Kansas.

Lewis, Charlton T. & Charles Short. 1879. *A Latin dictionary: Founded on Andrews' edition of Freund's Latin Dictionary: Revised, enlarged, and in great part rewritten by Charlton T. Lewis, Ph.D, and Charles Short, LL.D.* Oxford: Clarendon Press. http://www.perseus.tufts.edu

Marantz, Alec. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *Penn Linguistics Colloquium (PLC)* 21. 201–225.

Marantz, Alec. 2001. Words (ms.). Cambridge, MA: MIT. http://babel.ucsc.edu/~hank/mrg.readings/Marantz_words.pdf

Marchand, Hans. 1969. *The categories and types of present-day English word-formation*. Munich: C.H. Beck.

Maynor, Natalie. 1979. The morpheme *un*. *American Speech* 54(4). 310–311. https://doi.org/10.2307/454704

Mendívil-Giró, José Luis. 2015. *El cambio lingüístico: sus causas, mecanismos y consecuencias*. Madrid: Síntesis.

Mendívil-Giró, José Luís. 2016. ¿Qué relación hay entre el cambio lingüístico y la evolución del lenguaje? *Linred: Lingüística en la Red* 14.

Montero Curiel, María Luisa. 1999. *La prefijación negativa en español*. Cáceres: Universidad de Extremadura.

Newell, Heather. 2008. Aspects of the morphology and phonology of phases. Montreal, Canada: McGill University dissertation. https://escholarship.mcgill.ca/concern/theses/1c18dh59v

Oltra-Massuet, Isabel. 2014. *Deverbal adjectives at the interface: A crosslinguistic investigation into the morphology, syntax and semantics of -ble*. Berlin: Mouton de Gruyter (Interface Explorations 28).

Pantcheva, Marina. 2011. Decomposing path: The Nanosyntax of directional expressions. Tromsø, Norway: University of Tromsø dissertation. http://hdl.handle.net/10037/3631

Pinkster, Harm. 2015. *Oxford Latin syntax, Volume I: The simple clause*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199283613.001.0001

Ramchand, Gillian. 2008. *Verb meaning and the lexicon: A first phase syntax*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511486319

Selkirk, Elizabeth. 1982. *The Syntax of Words*. Cambridge, MA: MIT Press.

Siegel, Dorothy. 1974. Topics in English Morphology. Cambridge, MA: MIT dissertation.

Starke, Michal. 2009. A short primer to a new approach to language. *Nordlyd* 36(1). 1–6. https://doi.org/10.7557/12.213

Starke, Michal. 2014. Towards elegant parameters: Language variation reduces to the size of lexically stored trees. In M. Carme Picallo (ed.), *Linguistic variation in the Minimalist framework*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198702894.003.0007

Svenonius, Peter. 2010. Spatial P in English. In Guglielmo Cinque & Luigi Rizzi (eds.), *The cartography of syntactic structures*, Volume 6, 127–160. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195393675.003.0004

Tranel, Bernard. 1976. A Generative treatment of the prefix *in-* of modern French. *Language* 52(2). 345–369. https://doi.org/10.2307/412564

Varela, Soledad. 1983. Lindes entre morfemas: el prefijo negativo *in-*. In Fernando Lázaro Carreter & Emilio Alarcos Llorach (eds.), *Serta Philologica: F. Lázaro Carreter. Natalem diem sexagesimum celebranti dicata*, Volume 1, 637–648. Madrid: Cátedra.

Varela, Soledad. 1990. *Fundamentos de Morfología*. Madrid: Crítica.

Varela, Soledad. 2002. Active or subjective adjectival-participles in Spanish. In James F. Lee, Kimberly L. Geeslin & J. Clancy Clements (eds.), *Structure, meaning, and acquisition in Spanish. Papers from the 4th Hispanic Linguistics Symposium*, 304–316. Somerville, MA: Cascadilla Press.

Varela, Soledad & Josefa Martín García. 1999. La prefijación. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, Volume 3, 4993–5040. Madrid: Espasa-Calpe.

Zimmer, Karl E. 1964. *Affixal negation in English and other languages: An investigation of restricted productivity*. Supplement to *Word* 20(2), Monograph 5.

Code-mixing and semantico-pragmatic resources in francophone Maine

Meanings-in-use of *yeah/yes* and *ouais/oui*

Kendall Vogh
York University

This chapter reports some results of an exploratory corpus study (Vogh 2018) investigating whether bilingual speakers might use code-mixing to leverage the contextualized meanings (i.e., meanings-in-use) of specific lexical resources that happen to be 'in the other code'. A total of 206 code-mixed tokens of *yeah, yes, ouais*, and *oui* from nine videotaped oral history interviews of Franco-Americans in Maine are considered. Drawing on sociolinguistic, qualitative semantic, and discourse-analytic approaches, I find that the speakers studied do prefer different meanings-in-use for resources from their different languages, suggesting that "what speakers wish to say" (Backus 2001: 150) is indeed a relevant factor in code-mixing and in how bilingual speakers experience their bilingualism.

Keywords: code-mixing (code-switching), contextualized lexical semantics, meanings-in-use, bilingual speech, Franco-Americans, French in Maine

1. Introduction

Code-mixing – broadly, the juxtaposition of elements from more than one linguistic code in the same utterance, conversation, or text – has been studied from a number of different perspectives, particularly with regard to how its occurrence is conditioned or constrained by syntactic, social, and interactional factors (see Gardner-Chloros 2009 for a review). However, as Backus (2001: 150) points out, code-mixing "is not just determined by what is syntactically possible, but also by *what speakers wish to say*" (my emphasis), in other words, by the semantico-pragmatic meanings bilingual (or multilingual) speakers might wish to express. It is well-known that different languages lexicalize meanings differently; for example, the meanings expressed by the distinct English lexical units *like* and *love* are both expressed in French by the single lexical unit *aimer*. Common sense suggests that a bilingual speaker who has all of these lexical resources at their disposal, given an interactional context where

code-mixing is an acceptable practice, could avail themselves of whichever item most closely or directly expresses their referential intent, changing code to do so if necessary.[1] In spite of this, the semantico-pragmatic component of code-mixing has only rarely been studied. Focused documentary research identified only two directly relevant studies: Backus (2001) finds that his Dutch- and Turkish-speaking participants are more likely to code-mix words with high semantico-pragmatic specificity, while King & Nadasdi (1999) conclude that expressions like *I think* and *I guess* are used by Newfoundland francophones due to the higher degree of uncertainty they express versus similar expressions in French.

This paper reports some results of a larger study (Vogh 2018) which sought to expand this limited literature on code-mixing as a strategy of semantico-pragmatic expression, that is, on whether bilingual speakers might use code-mixing to leverage the contextualized meanings (i.e., meanings-in-use) of specific lexical resources that happen to be 'in the other code'. As so little research exists in the area, the primary goal of this study was to develop and apply a methodology for studying the semantico-pragmatic dimension of code-mixed items. The corpus study was exploratory, so the resulting data sets were relatively small, but this paper presents the results of the most frequently occurring code-mixed lexical unit in the corpus, *yeah*, as well as *yes*, supplemented with a comparison to the French units *ouais* and *oui*.

2.   Approach and key terms

2.1   Code and code-mixing

The study of code-mixing forces the researcher to confront some of the most fundamental questions in linguistics in that, to identify instances where speakers are using more than one language, one must take some position on what constitutes 'a language' and 'what counts as different languages'. This paper takes the theoretical stance that named languages are not inherently empirical objects, but rather socially constructed codes. Specifically, what we consider to be a language (or for that matter a dialect, or a register) is in fact a collection of semiotic practices, interpreted as a meaningful and distinct entity by those who use or encounter it (Otheguy, García & Reid 2015). This does not, however, imply that named languages are any less 'real' in terms of their power to structure understanding and influence actions.

The principal consequences of this stance for the present study are, first, that "a bilingual speaker [does] not possess two language systems but one: their repertoire of linguistic resources" (Vogh 2018: 9, my translation; see also Franceschini 1998; Lüdi & Py 2009). Thus, second, code-mixing is conceptualized not as 'switching' between language systems which produce empirically distinct outputs, but as

---

1.   This paper uses singular *they* as a gender-neutral third-person pronoun.

the selection of resources from the speaker's repertoire that are associated, by the speaker and others, with different socially meaningful codes (Franceschini 1998; Otheguy, García & Reid 2015). Third, it is participants' associations of resources to codes, which may vary according to the social and interactional norms in play, that determine what constitutes a change in code (cf. Auer 1995).

This last point is consequential with regard to the question of borrowing. The distinction between borrowing and code-mixing figures prominently in the literature (e.g. Pfaff 1979; Poplack & Meechan 1998), especially in the case of single lexical items as in the present study, and with good reason: if a lexical unit is judged to be borrowed, then its use does not, in fact, constitute a change in code. However, what is being considered in the present analysis is not whether a unit 'belongs' to a given code phonetically, grammatically, or historically, but whether it is associated with that code according to the social norms to which the speakers in the corpus hold. Hence, although some authors have explicitly treated *yeah* as a borrowing in other francophone minority communities (e.g., Walker 2005 for Albertan French),[2] in the present study *yeah* and *yes* were found to be so strongly associated with the English code that their use could be considered a code change from French.[3]

Finally, the terminology one uses also ultimately points to certain theoretical and methodological choices. I have opted to use the term 'code-mixing', rather than, for example, 'code-switching' or 'translanguaging' (cf. Otheguy et al. 2015; Wei 2018) in this study, to represent the phenomenon according to what speakers in the corpus appear to be doing – combining resources from distinct socially-recognized language codes – rather than how or why they may be doing it.

---

2. In fact, there is a substantial literature on borrowed discourse markers in a number of speech communities and across a number of different languages (e.g. Brody 1995; Fuller 2001; King 2008). That discourse markers are so frequently the objects of borrowing or code-mixing is thought to be due to their efficiency at communicating complex, nuanced pragmatic information; their high frequency in discourse; and their pragmatic and syntactic detachability (Dostie 2004; Matras 1998; Vogh 2018). However, it should be noted that there is a longstanding tradition in code-switching research of treating single-word switches ipso facto as borrowings, particularly if they are frequently occurring in the corpus under study (e.g. Poplack, Sankoff, & Miller 1988). Discourse markers are typically both single words (or complex lexical items that function as a single unit), and frequent words in spontaneous speech data. The ubiquity of the above assumptions therefore suggests that discourse markers are overrepresented in the borrowing literature, and underrepresented in the code-mixing literature. Indeed, I suspect that in many cases where other-language discourse markers have previously been analyzed as borrowings, an argument could be made (as I have done here) to instead treat them as instances of code-mixing. Given this, and given the present study's focus on elaborating methods for studying code-mixing, I do not draw further on the discourse-marker-borrowing literature in this chapter.

3. This decision was made using seven criteria pertaining to the frequency, distribution, and integration of a given unit within the corpus and in the wider linguistic community (see Vogh 2018, § 4.1.4 for details).

246 Kendall Vogh

2.2   Meaning-in-use

Much of what individual speakers know of meaning is not explicitly or formally taught, but is learned through experience, inference, and at times negotiation with other speakers; that is, meaning in general is situated and highly contextualized. This means that, in studying lexical meaning in conversation, we must observe the 'meaning-in-use' (*sens en usage*, see along these lines Courbon 2015): the meaning conveyed to interlocutors – and eventually, at least in part, to the analyst – by a particular lexical unit, used by a particular speaker, in a particular (conversational, interpersonal, historical, sociopolitical, etc.) context. Considering meaning in this way necessarily collapses the distinction between the semantic and the pragmatic – in the present analysis, between meaning and discursive function – as both are part of the conceptual reality that speakers are using the lexical unit to refer to, and part of what interlocutors make use of in interpreting that act of reference.[4]

With these theoretical and terminological points in mind, I turn now to the study itself, beginning with a discussion of the corpus and speakers under analysis.

3.   Corpus and context

3.1   The corpus

The data presented in this paper come from a corpus of audio-visual recordings of loosely structured conversation between French-English bilinguals residing in Maine, in the United States. This corpus (one of three subcorpora used in the larger study) was constituted from the video archives of the oral history project of the Franco-American Centre of the University of Maine, Orono, used with permission. The interviews thus focus on the interviewee's lived experiences as both a francophone and a resident of (rural) Maine, particularly in the early- to mid-20th century. In order to constitute an appropriately sized corpus for a non-automated, exploratory analysis, a 50% sample was selected from among those interviews in the archive (1) recorded between 2001 and 2012, and (2) whose metadata indicated the use of both English and French. Care was taken to keep the sample proportional with regard to available metadata: gender and number of participants and length of recording. One video from the sample is not reported on here, as its speakers differ

---

4.   I here follow those traditions in which 'reference' is understood as the act of indicating a particular reality –,physical, mental, emotional, etc. –,as the speaker conceives of and orients to it, i.e., a conceptual reality. For example, both *that dress* and *those rags* may describe the same physical object, but present different conceptual realities, and thus constitute different acts of reference.

from those of the other 8 on certain key traits (region of origin, variety of French, social networks) that may influence their code-mixing practices. The final corpus under analysis in the present paper thus consists of 8 video interviews totaling 13 recorded hours.

A total of 12 speakers contributed to the corpus, in constellations varying between two and four participants at any one time. The eight videos in the corpus all have one participant in common: the community member responsible for the bulk of the oral history project, who acts as the discussion facilitator and is also a friend or acquaintance of his interviewees. Overall, the corpus was found to contain 2,691 instances of code-mixing, or 207 instances per recorded hour.

3.2   The context: Franco-Americans of Maine

Because the study consists of a secondary analysis of data originally collected for another purpose, no systematically-collected personal or demographic information about the speakers exists – only what they choose to reveal in their interviews. The available information thus is not comparable for all speakers in the corpus; however, the speakers do all share an identity as Franco-Americans. The brief sketch of the Franco-Americans of Maine that follows may therefore help to contextualize some of the speakers' (linguistic) experiences and choices.[5]

Recent estimates suggest that roughly 24% of the population of Maine identifies as Franco-American, making them the largest single ethnic group in the state. Within this group, surveys suggest that 28% consider themselves fluent in French (Albert et al. 2013). Broadly, Franco-Americans in Maine can be divided into two subgroups: descendants of Acadians who settled the rural St. John River valley (near present-day New Brunswick) in the 18th century, who speak an Acadian variety of French; and descendants of Québécois migrants who settled primarily in southern industrial towns in the 19th and early 20th centuries, who speak a Laurentian variety of French (Allen 1974).

Historically, Franco-Americans have faced discrimination and social and economic exclusion. They worked primarily as unskilled, low-paid labourers (farmers, loggers, and mill and factory workers) and often had limited English proficiency and low educational attainment, giving rise to damaging 'dumb Frenchmen' stereotypes (Albert et al. 2013; Allen 1974). Amid the increased xenophobia of wartime and postwar America, their language and Catholic religion made them targets for increased discrimination and violence, including attacks by the Ku Klux Klan

---

5.   For further discussion on the subject, I refer the reader to Fox's work on French and Franco-Americans in New England, e.g., Fox (2007).

(Richard 2009). Languages other than English were prohibited in Maine public schools in 1919; students received corporal and psychosocial punishments for using French on school grounds (Albert et al. 2013; Allen 1974). This law was not repealed until 1969 and then only to allow transitional bilingual education. Although French remained a language of instruction in parochial schools in francophone communities – an important factor in French-language maintenance and the acquisition of French literacy – there was little support, institutional or otherwise, for French outside of the francophone community. Wherever Maine francophones came into contact with the non-francophone world, they faced not only pressure to speak English, but also pressure *not to be speakers of French*. Many of Albert et al. (2013)'s respondents report that they or their relatives deliberately refused to transmit French to future generations in order to spare them the social stigma of not being native English speakers. In spite of this, as the census and survey data suggest, French still survives in Maine today to a surprising extent, although with vanishingly few young speakers, its future is gravely threatened.

The speakers in the corpus reflect elements of this history. All are from the Saint John Valley, and all are at least 60 years old. All are fluent French and English speakers, having learned French as a first or home language and used principally English at school, at work, and in the wider community. Many of their descendants, meanwhile, never learned or no longer speak French. Some speakers mention having little ability to read or write in French due to Maine's English-only education policies. Some mention facing discrimination or rejection for speaking French; one recalls witnessing an anti-francophone Ku Klux Klan demonstration. All speakers readily mix French and English in conversation with other in-group members; although some are aware that such mixing is dispreferred in 'standard' North American French, they also describe it as being typical of the local variety, which they seem to prefer.

4. Data and coding

4.1 Identifying and selecting occurrences of code-mixing

The 2,691 occurrences of code-mixing identified in the corpus include any juxtaposition of elements representing linguistic codes recognized as different languages by the participants (to the extent that participants' perceptions could be determined – see § 2.1). This includes code changes of any length, in any direction (from French to English or from English to French), and in any position (within the same utterance, between utterances, or between speakers). However, only isolatable lexical items (*n* = 1,325; 931 English and 394 French) were retained for further analysis.

A lexical unit was qualified as isolatable if it was either: (1) a single other-language lexical item in a given stretch of talk, or (2) the first lexical unit in a longer stretch of talk in the 'other' code, provided it was syntactically independent from the rest of that stretch. These conditions increase the likelihood that the meaning-in-use of the lexical unit in question was the reason behind the change in code.

Due to the prevalence in the larger study of English code-mixed items, these formed the basis of the analysis, and the analysis of French-language items was limited to a comparison with the selected English units. Since it is rare in naturalistic code-mixing data for any single lexical item to recur with sufficient frequency for analysis – for example, only 11 individual English lexical items occur 10 times or more in the corpus, and only four of these occur 20 times or more – the isolatable occurrences were lemmatized and related lemmas grouped together, in order to allow usage patterns to emerge.[6] *Yeah/yes* was the most frequently occurring such group in the corpus; isolatable occurrences of *ouais* and *oui* form the French-associated comparison group.[7] Table 1 indicates the frequency of occurrence of each of these units in the corpus, separated by speaker (speakers that did not produce any occurrences are not shown).

Table 1. Frequency of occurrence of isolatable tokens of *yeah, yes, ouais* and *oui* by speaker

Speaker*	Age and gender	No. of isolatable tokens				Total per speaker (% of all tokens)
		yeah	*yes*	*ouais*	*oui*	
FAC 01–08: AN	63–74 M	56	2	16	5	79 (38.3%)
FAC 08: INT	61 M	25	8	5	8	46 (22.3%)
FAC 06: INT1	84 M	24	1	6	0	31 (15.0%)
FAC 01: INT	81 F	16	3	6	5	30 (14.6%)
FAC 03: INT1	80 M	12	1	0	1	14 (6.8%)
FAC 04: INT1	95 M	2	0	0	0	2 (1.0%)
FAC 06: INT2	64 M	2	0	0	0	2 (1.0%)
FAC 07: INT	77 M	1	0	0	1	2 (1.0%)
Total tokens		138	15	33	20	206
		153		53		

* Speakers are identified by corpus, recording number, and speaker role (AN = *animateur* or facilitator, INT = interviewee). A number is used to distinguish different interviewees from the same recording.

---

6. In descending order, the 11 highest-frequency items are: *yeah, high school*, the tag question *is that right?*, the exclamation *holy boy!*, *yes, so, band* (i.e., a musical act), *now, right, that's right*, and *anyway*.

7. *Yep* was also included in this category in the larger study, but due to the small number of isolatable tokens (*n* = 4) has been excluded from the present analysis. The comparable French unit *ouaip* had no isolatable tokens.

4.2   Coding of the selected occurrences: Categories of meaning-in-use for *yeah* and *yes*

The tokens under study were coded according to a number of categories of meaning-in-use, based on my observations while exploring the corpus data and informed by a review of the literature on both first- and second-language uses of *yeah* and *yes* (e.g., Drummond & Hopper 1993; House 2013; see Vogh 2018 for further discussion). A total of six categories of meaning-in-use were identified, presented in Table 2.

Table 2. Categories of meaning-in-use for *yeah/yes* and *ouais/oui*

Category Name	Description	Example*
Affirmative response	Gives assent or provides factual confirmation. Subdivided into responses to *explicit* or *implicit* (inferred) questions.	INT1: My mother could make thirteen different types of: patates fricassées! [*potatoes fricasee*] AN: [laughs] Vous en avez mangé, des patates! < INT1: yeah! > (FAC 03, 16:19)
Alignment	Indicates understanding, agreement, or sympathy.	INT: [*of his time at boarding school:*] pis le pain était tout le temps frais. Y faisait du pain eux autres là [*and the bread was always fresh. They made it themselves*] < AN: c'est vrai? [*is that so ?*]> la journée ouais [*every day, yeah*] < AN: yeah > (FAC 07, 48:57)
Emphatic expression	Indicates excitement, enthusiasm, or other emotional response. Includes emphatic repetition.	AN: Vous mangez ben par exemple. [*you ate well, though*] < INT: oh yes! oh yes. > Uh that's what I hear, most people didn't know they were poor. (FAC 03, 11:10)
Continuer	Indicates attentive listening and encourages the interlocutor to continue their turn.	INT: I think we were a hundred and twenty. M'en rappelle plus là. [*I don't remember anymore now*] < AN: yeah > Was (.) a hundred and (.) fairly big class (FAC 08, 25:36)
Subject change	Closes a current conversation topic and/or opens a new one.	INT: moi j'écrivais ça dans le papier ça fait qu'ils voulaient pas que ça paraisse dans le papier! [laughs] [*I was writing all this in the paper, so they didn't want it to appear in the paper!*] AN: ah: yeah (.) Were you there when Father *** was there? (FAC 01, 1:24:37)
Repair	'Smooths over' a disfluency or other conversational trouble spot. Subdivided into *self-oriented repair* or *facilitation* (other-oriented repair).	INT: y faisait chaud l'ét - en l'été hein? [*It was hot in the summer, eh ?*] < INT2: yeah > je prenais ça. Je mettais la glace! [*I would take this. I would put ice in it!*] (FAC 06, 1:00:27)

* Excerpts are identified by corpus name, recording number, and start time of the excerpt. Angle brackets indicate speech that does not affect overall possession of the floor (e.g., backchannel). Prolonged syllables are indicated by ":" and pauses by "(.)". English glosses are provided in italics within square brackets immediately following the French utterance.

These categories are not mutually exclusive, and often overlap – indeed, pragmatic markers like *yeah* are especially useful to speakers precisely because they can express many related meanings at the same time. Each isolatable occurrence of *yeah* and *yes* was thus coded with all of the categories that applied to that instance, and the same coding procedure was then repeated for *ouais* and *oui*.

5.   Results and interpretation

Table 3 presents the absolute (#) and relative (%) frequencies with which each category and subcategory of meaning-in-use was attributed to an occurrence of the English and French lexical unit groups.[8] Given the imbalance in the total number of occurrences of each language group (153 versus 53), relative frequency was calculated in terms of the number of units of the same group. For example, 81.7% of the time that either *yeah* or *yes* is uttered as an isolatable change in code, it is used as an affirmative response.

It should be noted that all four units were used with nearly all of the meaning-in-use categories (*yes* was not used for self-repair nor as a continuer, and *oui* was not used for self-repair). In other words, the key difference between *yeah/yes* and *ouais/oui* is not *what* speakers are using these units to do in discourse, but *how often* they do so. Table 3 shows that for the 'canonical' meaning of these units, as an affirmative response to an explicit question, the speakers have no strong preference for English or French. French units are slightly preferred for implicit/inferred questions, and to express alignment. Stronger preferences for the French units can be observed for emphatic expression, use as a continuer, and repair, particularly self-repair. The English units, meanwhile, are somewhat preferred for self-repair and for subject changes.

Overall, what Table 3 shows is that the French units are preferred, to varying degrees, for those meanings-in-use most associated with (inter)subjectivity. Responding to an implicit question requires attending to the interlocutor's interactional habits and prior knowledge to ascertain whether a response is in fact needed. Alignment requires understanding and sympathizing with another's emotional state. Continuers, too, require attention to the interlocutor's state of mind, for example, whether they might want or need validation before continuing their turn. Emphatic expression concerns both one's own emotional state and that of the interlocutor, such as whether they are hoping for a particularly enthusiastic

---

8.   While some within-language differences were noted, these were unsurprising (e.g., *yes* and *oui* are used relatively more often to respond to explicit questions than *yeah* and *ouais*, respectively) and will not be discussed further here.

Table 3. Frequency of meaning-in-use by language

Category	Affirmative response		Alignment		Emphatic	
	#	%	#	%	#	%
yeah / yes (*n* = 153)	125	81.7	91	59.5	57	37.3
ouais / oui (*n* = 53)	47	88.7	37	69.8	30	56.6
Highest language group (% difference)		FR (7.0)		FR (4.1)		FR (19.3)

Subcategories	*Explicit question*		*Implicit question*		*N/A*	*N/A*
	#	%	#	%		
yeah / yes (*n* = 153)	40	26.1	85	55.6		
ouais / oui (*n* = 53)	15	28.3	32	60.4		
Highest language group (% difference)		EN (0.7)		FR (4.9)		

Category	Continuer		Subject change		Repair	
	#	%	#	%	#	%
yeah / yes (*n* = 153)	54	35.3	42	27.5	56	36.6
ouais / oui (*n* = 53)	30	56.6	12	22.6	29	54.7
Highest language group (% difference)		FR (21.2)		EN (4.9)		FR (18.1)

Subcategories	*N/A*		*N/A*		*Self-Repair*		*Facilitation*	
					#	%	#	%
yeah / yes (*n* = 153)					7	4.6	19	12.4
ouais / oui (*n* = 53)					1	1.2	16	30.2
Highest language group (% difference)						EN (3.4)		FR (17.8)

response. Similarly, facilitation requires empathizing with the interlocutor's inter-actional trouble (a lexical search, a difficult topic, etc.) and helping them continue the interaction. On the other hand, the categories where the English lexical units were preferred are mainly associated with the 'business' of conducting conversation: correcting one's own conversational misfires (self-repair) and managing the topic of conversation (subject change).

This finding makes sense given the sociohistorical context of Maine franco-phones. They have suffered violence and ostracization because of their language. Their French proficiency both unites them as a group and sets them apart from other Mainers, even in many cases from their own children and grandchildren. In other words, the only people with whom they have many of their significant expe-riences, knowledge, and practices in common are other francophone Mainers. It makes sense, then, that in conversation with each other, the Maine francophones in the corpus have come to prefer French-associated linguistic resources for those functions which represent and reinforce social cohesion, personal connections, and the sense of belonging, comfort, safety, and freedom of expression that comes from being *entre nous*.

It should be noted that these results differ somewhat from the results for other lexical units in Vogh (2018). Notably, in the case of the pragmatic marker *so* and the expression *is that right?*, it appears to be the English lexical units that are preferred over the comparable French items (*alors/ça fait que* and *c'est[-tu] vrai?* respectively). However, in both of these cases, there are noticeable differences in the meanings-in-use for the English versus the French lexical items. For example, *so* appears to have more versatility in meaning than *ça fait que*, which could explain why it occurs in more contexts. In the present case, as previously stated, there is no real difference in meanings-in-use for *yeah* and *yes* compared to *ouais* and *oui*. Instead, what appears operational in the case of *yeah/yes* and *ouais/oui* is a contextual (in both the broad and narrow sense) influence on speaker preferences.

6.   Conclusion

This paper has reviewed the results of an exploratory corpus study investigating how bilingual speakers make use of the semantico-pragmatic resources in their linguistic repertoire through code-mixing. We have seen that *yeah* and *yes*, when the focus of code-mixing, do not differ from *ouais* and *oui* in terms of the types of meaning-in-use they can potentially express; instead, they differ in terms of which lexical units speakers in the corpus appear to prefer for expressing those meanings, with French-associated units being preferred for meanings related to

(inter)subjectivity and interpersonal dynamics. This preference may be related to the significance of French as a marker of in-group identity and shared experience.

The analysis presented in this paper has certain limitations. Most notably, since the larger study on which it is based was exploratory in nature, the dataset under analysis is small (a total of 206 tokens). This both precludes any analysis of statistical significance and increases the relative effect of the usage habits of any particularly prolific speaker on the results, such as the principal discussion facilitator (see Table 1). Additionally, the results are the work of a single coder (the author); as all qualitative work involves an element of subjectivity in the analysis, a different coder might have arrived at somewhat different results.

In spite of these limitations, the results indicate that bilingual speakers do make use of resources associated with their different languages to relatively different referential effects, suggesting that 'what speakers wish to say' is indeed a relevant factor in code-mixing. Code-mixing allows speakers to leverage specific semantico-pragmatic resources when it is socially and interactionally acceptable to do so, and thus functions as a resource in itself. Moreover, the results indicate that the overall goal of the corpus study, to develop a methodology that can account for the semantico-pragmatic factors involved in code-mixing, has been achieved. Future work will focus on continuing to refine that methodology (e.g., refining the measure of ambiguity, recruiting additional coders and incorporating measures of intra- and inter coder reliability) and applying it to more and different data (e.g., lexical items in non-isolatable positions, a larger sample from the Franco-American Centre archives, additional corpora of bilingual conversation).

In sum, the referential component of code-mixing absolutely deserves further study, and not only in the interests of descriptive completeness, observing language variation and change in contact situations, or as a way to test theories of language acquisition or bilingual speech production. Code-mixing as a means of more fully or precisely expressing one's referential intentions is part and parcel of the linguistic practices and communicative competencies that make up the daily lives of bi- and multilingual speakers around the world. In drawing attention to this fact, this study joins a growing body of research across linguistic disciplines calling for theories, approaches, data sets and methodologies that can account for, normalize, and contextualize these practices.

Acknowledgements

Funding

References

Albert, Jacob, Tony Brinkley, Yvon Labbé & Christian Potholm. 2013. *Contemporary attitudes of Maine's Franco-Americans* (Franco-American Centre Occasional Papers 1). Orono, ME: Franco-American Centre Publications.

Allen, James P. 1974. Franco-Americans in Maine: A geographical perspective. *Acadiensis* 4(1). 32–66.

Auer, Peter. 1995. The pragmatics of code-switching: A sequential approach. In Lesley Milroy & Pieter Muyksen (eds.), *One speaker, two languages: Cross-disciplinary perspectives on code-switching*, 115–135. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620867.006

Backus, Ad. 2001. The role of semantic specificity in insertional codeswitching: Evidence from Dutch-Turkish. In Rodolfo Jacobson (ed.), *Codeswitching worldwide, II* (Trends in Linguistics: Studies and Monographs 126), 125–154. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110808742.125

Brody, Jill. 1995. Lending the 'unborrowable': Spanish discourse markers in indigenous American languages. In Carmen Silva-Corvalán (ed.), *Spanish on four continents: Studies in language contact and bilingualism*, 132–147. Washington, D.C.: Georgetown University Press.

Courbon, Bruno. 2015. Pratiques sémantiques et différences interindividuelles à l'ère des corpus informatisés [Semantic practices and interindividual differences in the era of computerized corpora]. *Les Cahiers de lexicologie* 106(1). 91–126.

Dostie, Gaétane. 2004. *Pragmaticalisation et marqueurs discursifs : Analyse sémantique et traitement lexicographique* [Pragmaticalization and discourse markers: Semantic analysis and lexicographical treatment]. Brussels: De Boeck Duculot. https://doi.org/10.3917/dbu.dosti.2004.01

Drummond, Kent & Robert Hopper. 1993. Some uses of *yeah*. *Research on Language and Social Interaction* 26(2). 203–212. https://doi.org/10.1207/s15327973rlsi2602_6

Fox, Cynthia A. 2007. Franco-American voices: French in the northeastern United States today. *The French Review* 80(6). 1278–1292.

Franceschini, Rita. 1998. Code-switching and the notion of code in linguistics: Proposals for a dual focus model. In Peter Auer (ed.), *Code-switching in conversation: Language, interaction and identity*, 51–72. London: Routledge.

Fuller, Janet M. 2001. The principle of pragmatic detachability in borrowing: English-origin discourse markers in Pennsylvania German. *Linguistics* 39(2). 351–369. https://doi.org/10.1515/ling.2001.014

Gardner-Chloros, Penelope. 2009. *Code-Switching*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511609787

House, Juliane. 2013. Developing pragmatic competence in English as a Lingua Franca: Using discourse markers to express (inter)subjectivity and connectivity. *Journal of Pragmatics* 59A. 57–67. https://doi.org/10.1016/j.pragma.2013.03.001

King, Ruth. 2008. *Chiac* in context: Overview and evaluation of Acadie's *joual*. In M. Meyerhoff & N. Nagy (Eds.), *Social lives in language – Sociolinguistics and multilingual speech communities: Celebrating the work of Gillian Sankoff*, 137–178. Amsterdam: John Benjamins. https://doi.org/10.1075/impact.24.12kin

King, Ruth & Terry Nadasdi. 1999. The expression of evidentiality in French-English bilingual discourse. *Language in Society* 28. 355–365. https://doi.org/10.1017/S0047404599003024

Lüdi, Georges & Bernard Py. 2009. To be or not to be … a plurilingual speaker. *International Journal of Multilingualism* 6(2). 154–167. https://doi.org/10.1080/14790710902846715

Matras, Yaron. 1998. Utterance modifiers and universals of grammatical borrowing. *Linguistics* 36(2). 281–331. https://doi.org/10.1515/ling.1998.36.2.281

Otheguy, Ricardo, Ofelia García & Wallis Reid. 2015. Clarifying translanguaging and deconstructing named languages: A perspective from linguistics. *Applied Linguistics Review* 6(3). 281–307. https://doi.org/10.1515/applirev-2015-0014

Pfaff, Carol W. 1979. Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language* 55(2). 291–318. https://doi.org/10.2307/412586

Poplack, Shana, & Marjory Meechan. 1998. Introduction: How languages fit together in code-mixing. *International Journal of Bilingualism* 2(2). 127–138. https://doi.org/10.1177/136700699800200201

Poplack, Shana, David Sankoff, & Christopher Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26. 47–104. https://doi.org/10.1515/ling.1988.26.1.47

Richard, Mark Paul. 2009. "This is not a Catholic country": The Ku Klux Klan confronts Franco-Americans in Maine. *New England Quarterly* 82(2). 285–303. https://doi.org/10.1162/tneq.2009.82.2.285

Vogh, Kendall. 2018. Ressources linguistiques et visée référentielle chez des individus bilingues français-anglais: L'alternance codique comme stratégie d'expression sur le plan lexical [Linguistic resources and referential intentions in French-English bilinguals: Code-switching as a strategy of lexical expression]. Québec, Canada: Université Laval MA thesis. http://hdl.handle.net/20.500.11794/29802

Walker, Douglas. 2005. Le français dans l'Ouest canadien [French in the Canadian West]. In Albert Valdman, Julie Auger & Deborah Piston-Hatlen (eds.), *Le français en Amérique du Nord: État présent* (Langue française en Amérique du Nord), 187–205. Quebec City, Canada: Presses de l'Université Laval.

Wei, Li. 2018. Translanguaging as a practical theory of language. *Applied Linguistics* 39(1). 9–30. https://doi.org/10.1093/applin/amx039

CHAPTER 15

Exceptionality and ungrammaticality in Spanish stress
A Stratal OT approach

Katerina Tetzloff

University of Massachusetts Amherst

Regular Spanish stress is stem-final, but many exceptions exist. Though exceptions must be accounted for, they must not yield ungrammaticality. The current proposal distinguishes exceptionality from ungrammaticality by arguing for some degree of quantity sensitivity in Spanish stress assignment. Past approaches that do not take syllable weight into account predict that a nonce input, like *rapind-o*, if marked as exceptional, will bear ungrammatical, antepenultimate stress (*rápindo). This proposal resolves this by analyzing Spanish stress using Stratal OT, where Regular stress is stem-final. Lexically-indexed constraints allow for each of the exceptional stress patterns, without yielding ungrammatical forms. An exceptional nonce input like *rapind-o* appropriately receives penultimate stress (*rapíndo*) regardless of if it is marked as exceptional, thus not conflating exceptionality with ungrammaticality.

Keywords: phonology, stress, Spanish, Stratal OT, exceptionality

1. Introduction

There have been numerous past analyses of Spanish stress (Harris 1982, 1989; Roca 1988; Dunlap 1992; Lipski 1997; Baković 2016; Doner 2017), but a challenge that has not yet been met is to account for the fact that some Spanish stress generalizations allow exceptions, while others are exceptionless. This paper presents a novel account for these data that utilizes lexically-indexed constraints (Pater 2000, 2009) in a Stratal Optimality Theoretic (henceforth, Stratal OT) framework (Prince & Smolensky 1993/2004; Kiparsky 2000; Bermúdez-Otero 2008). This analysis argues for some degree of quantity sensitivity at the level of the morphological word-stem, allowing for the differentiation between exceptional versus ungrammatical stress assignment.

2.   Data

The position of primary stress varies across words in Spanish, but all stress assignment adheres to two inviolable and, therefore, exceptionless generalizations: first, primary stress in Spanish non-verbs[1] (henceforth, words) is strictly limited to a three-syllable window that is aligned to the right word edge and, second, words with heavy penults can never bear antepenultimate stress.[2] Beyond this, words tend to have stem-final stress (Regular pattern), which accounts for 95% of stress in Spanish words (Hualde & Nadeu 2014).

A broadly held view of Spanish morphology is that there are two types of suffixes: stem-level and word-level (Harris 1982; Roca 1988, among others). The stem-level suffixes include derivational morphology, whereas the word-level suffixes are inflectional, marking grammatical gender and number. These word-level suffixes are called terminal elements (TEs) (Harris 1982).

The word-stem is defined as the complete, morphologicallly-derived word excluding the TE (i.e., the root plus any additional stem-level morphemes). For example, if the diminuitive suffix, which is derivational, is suffixed to a root, this is at the stem-level. This suffix, along with the root, now forms part of the word-stem, to which the TE will attach, as in (1).

(1)   Table 1. Morphologically derived stems + TEs

stem	-TE	gloss	stem	-diminutive	-TE	translation
hermán	*-a*	"sister"	*herman*	*-ít*	*-a*	"sister (dim)"
lugár	*-ø*	"place"	*lugar*	*-cít*	*-o*	"place (dim)"
dominó	*-ø*	"domino"	*domino*	*-cít*	*-o*	"domino (dim)"

Regular stress in Spanish is stem-final (Roca 1988; Baković 2016), but there are multiple types of exceptional stress. Exceptional stress can appear one syllable to the left (Class 1), one syllable to the right (Class 2), or two syllables to the left (Class 3) of the Regular pattern. Examples of these regular and irregular stress patterns are shown in (2).

---

1.   Verb stress is lexical and contrastive between different tense, aspect, and mood suffixes. Because it behaves regularly and has no exceptions, further discussion of verbal stress assignment will be omitted from this paper.

2.   Spanish has adapted English loanwords that have heavy yet unstressed penults (e.g., *básketbol* "basketball"), where English stress is preserved, contra this generalization. Thus, it should be clarified that outputs of this form are unattested and ungrammatical in the Spanish lexicon but seem to be acceptable in loanwords.

(2) Table 2. Types of Spanish word stress

	Regular	**Class 1**	**Class 2**	**Class 3**
V-final	*hermán-a* "sister"	*rápid-o* "fast"	*dominó-ø* "domino"	--------------
C-final	*lugár-ø* "place"	*césped-ø* "sidewalk"	--------------	*régimen-ø* "diet"

In recognizing these three exceptional classes of Spanish stress, it is clear that a complete analysis of Spanish stress must be able to correctly predict each pattern yet still be restrictive enough to adhere to the two inviolable generalizations; this is a task that has not yet been met. Although past work has aimed to address both Regular and Class 1 exceptions, this has come at the cost of predicting ungrammatical forms. The goal of this paper is to present a novel account of Spanish stress assignment that not only can account for the Regular and all three exceptional classes of words but can also ensure that ungrammaticality is never predicted or conflated with exceptionality.

3.   Current proposal

This paper presents an account of Spanish stress assignment using Stratal OT and lexically-indexed constraints (Pater 2000, 2009). This analysis yields a unified account of both Regular and exceptional stress patterns and, crucially, never predicts an output that violates one of the inviolable Spanish stress generalizations. Furthermore, this analysis argues in favor of some degree of quantity sensitivity at the stem-level, which has been previously debated.

3.1   Stratal OT

In Stratal OT, phonology and morphology are stratified into separate stem, word, and post-lexical levels. Each stratum is subject only to its specific constraint ranking, which may differ in another stratum (Kiparsky 2000; Bermúdez-Otero 2008). Stems must satisfy the stem-level constraint ranking; words, the word-level ranking; and so on. The output of each stratum is the input to the next. Only stem- and word-level strata are used in the present analysis.

Stratal OT traditionally ranks the same constraints differently in different strata to account for opacity. Its purpose for the present analysis is two-fold. First, having separate stem-level and word-level strata is necessary because stress is assigned at the stem-level prior to TE suffixation. Second, the constraints used to assign stress in the first stratum are ranked differently in the second stratum. This allows the

position of stress to be maintained at the word-level when the TEs are added, so long as the additional morpheme and concomitant resyllabification do not cause the faithful stress to violate one of the two inviolable generalizations. If one of these generalizations is violated, the word-level constraints are ranked such that stress will be reassigned to conform to these generalizations.

3.2   Accounting for the Regular pattern

The first goal is to account for the Regular pattern of stem-final stress in Spanish. As previously put forward in the literature and described above, stems consist of morphologically-derived words, excluding TEs. For example, the stem of a word like *hermana* "sister" is *herman-*, and the stem of the diminutive form *hermanita* is *hermanit-*. To achieve the Regular pattern of stem-final stress, the following constraints are necessary:

> **WEIGHTBYPOSITION (WBP)**: Coda consonants are weight-bearing. Assign one violation for every coda consonant that is not associated with a mora.
>
> **FINALSTRESS**: Stress is stem-final. Assign one violation for every stem that does not bear stress on its final syllable (Baković 2016).[3,4]
>
> **NONFINALITY**: Final syllables should not be footed. Assign one violation for every final syllable that is footed (Prince & Smolensky 1993/2004).
>
> **FT-BINμ**: Feet are bimoraic. Assign one violation for every foot that does not consist of exactly two moras (McCarthy & Prince 1990).
>
> **TROCHEE**: Feet are trochaic. Assign one violation for every bisyllabic foot in which the rightmost syllable bears stress (McCarthy & Prince 1990).

In the present analysis, all coda consonants must be weight-bearing for reasons that will become apparent later on. To ensure this, WBP is undominated at the stem-level; therefore, it is omitted from the following tableaux, along with any candidates that violate this constraint.

Stem-final stress is primarily achieved by ranking FINALSTRESS above NONFINALITY, FT-BINμ, and TROCHEE. This is shown in the tableaux in (3)–(4) with words that both will and will not receive a TE, respectively.

---

3.   FINALSTRESS is categorical, contra Baković (2016). Multiple violations cannot be assigned when there is just one instance of the marked form.

4.   Categorical constraints are preferred over gradient constraints (i.e., constraints that assign multiple violations for one instance of the marked or unfaithful form) because gradient constraints are non-local and can make unattested predictions (Eisner 1997; McCarthy 2003).

(3) *moneda* "currency" stem-level

/moned-/	FINALSTRESS	FT-BINμ	NONFINALITY	TROCHEE
a. *(mó.)ned*	*!	*		
☞ b. *mo.(néd)*			*	
c. *(mo.néd)*		*!	*	*

(4) *lugar* "place" stem-level

/lugar-/	FINALSTRESS	FT-BINμ	NONFINALITY	TROCHEE
a. *(lu.gár)*		*!	*	*
☞ b. *lu.(gár)*			*	
c. *(lú.)gar*	*	*		

In (3), candidate (a) violates FINALSTRESS and FT-BINμ, while candidate (b) does not. Candidate (c) satisfies FINALSTRESS but violates the remaining constraints. Selecting the correct candidate with stem-final stress (b) is only possible if FINAL-STRESS outranks these two constraints. The same is true in (4): candidate (b) is the optimal because it satisfies FINALSTRESS and violates the fewest number of the lower ranked constraints compared to the other candidates.

As has been proposed by Harris (1982), among others, Class 2 exceptions are actually also Regular: the final vowel is part of the stem and not a TE. Thus, it is included in the stem-level input and receives final stress, as shown in (5).

(5) *dominó* "domino" stem-level

/domino-/	FINALSTRESS	FT-BINμ	NONFINALITY	TROCHEE
a. *do.mi.(nó)*		*!	*	
b. *(dó.mi)no*	*!			
☞ c. *do.(mi.nó)*			*	*

The winning stem-level output candidates serve as the input to the word-level stratum. If the word has a TE, it is attached here. As stress was assigned to the correct, stem-final syllable in the first stratum, the position of primary stress on the input needs to be maintained, despite any resyllabification that happens when TEs are added. The following additional constraint is thus necessary at this second stratum:

> HEADMATCH: In the output, stress falls on the same syllable as the input (i.e., the stem-level output). Assign one violation for every input segment that is stressed that does not correspond to an identical stressed segment in the output (McCarthy 2000).

In order to maintain the position of primary stress in this stratum, HEADMATCH must outrank any conflicting constraints that were active at the stem-level. Namely, this refers to FINALSTRESS, as Spanish stress is not word-final when a TE is present. The tableaux in (6)–(7) show the word-level constraint interactions for *moneda* "currency" and *lugar* "place".

(6)   *moneda* "currency" word-level

mo.(néd)+a	HEADMATCH	FT-BINμ	FINALSTRESS	NONFINALITY
a. *mo.(ne.dá)*	*!			*
☞ b. *mo.(né.da)*			*	*
c. *mo.(né.)da*		*!	*	

(7)   *lugar* "place" word-level

lu.(gár)+ø	HEADMATCH	FT-BINμ	FINALSTRESS	NONFINALITY
a. *(lu.gár)*		*!		*
☞ b. *lu.(gár)*				*
c. *(lú.)gar*	*	*!	*	

Candidate (a) in (6) shows that HEADMATCH must dominate FINALSTRESS at the word-level to have stress be maintained on the syllable it was assigned to at the stem-level. Evidence for crucially ranking FT-BINμ over NONFINALITY is shown by candidate (c). (7) shows the need to also rank HEADMATCH over NONFINALITY to ensure that the faithful candidate (b) is the winner.

3.3   Exceptional stress

The above constraints and rankings achieve the correct stress placement for Regular words in Spanish. However, this will incorrectly predict stress for Class 1 exceptions because of the fact that NONFINALITY applies variably at the stem-level, resulting in either stem-final or non-stem-final stress, which is then preserved at the word-level. Lexically-indexed constraints can achieve this type of exceptionality or variability. Lexically-indexed constraints (Pater 2000, 2009) are versions of general constraints that apply only to a specific class of indexed morphemes when the indexed morpheme overlaps with the locus of violation of the indexed constraint. For the present data, a lexically-indexed version of NONFINALITY, NONFINALITY$_1$, is needed.

   This constraint applies only to a specified class of morphemes.[5] Because this constraint picks out the final syllable of the stem as the locus of the constraint, the

---

5.   Not only root morphemes but also stem-level suffixes can bear exceptional indices. Regardless of the status of the root, if an indexed stem-level suffix is present, the word-stem is subject

indexed version is active when the indexed morpheme forms part of, or overlaps with, the final syllable of the word-stem. More specifically, this analysis follows Jarosz's (2018) proposal that the locality of the indexed constraint must be anchored to the right edge of the word-stem. Thus, indexed morphemes are subject to the lexically-indexed constraint if and only if the right edge of the structure referenced by the indexed constraint (i.e., the final syllable) is indexed as exceptional. For Spanish stress, Class 1 morphemes are those indexed to this lexically-specific constraint, defined below.

> NONFINALITY$_1$: Final syllables of stems indexed to this constraint must not be footed. Assign one violation for every indexed final syllable that is footed.

The goal of this constraint is to shift stress one syllable to the left of where it is predicted from the Regular pattern. It must outrank the general version of NONFINALITY, as well as both FT-BINµ and FINALSTRESS. (8)–(9) provide tableaux of an example of a Class 1 exception.

(8)   *rápido* "fast" stem-level

/rapid$_1$-/	NONFINALITY$_1$	FINALSTRESS	FT-BINµ	NONFINALITY
☞ a. *(rá.)pid*		*	*	
b. *ra.(píd)*	*!			*

(9)   *rápido* "fast" word-level

(rá.)pid + o	HEADMATCH	FINALSTRESS	NONFINALITY
☞ a. *(rá.pi.)do*		*!	
b. *ra.(pí.do)*	*!	*	*
c. *ra.(pi.dó)*	*!		*

To achieve exceptional stress assignment in (8), the lexically-indexed NONFINALITY$_1$ must outrank FINALSTRESS and FT-BINµ. Any candidate of a stem that is indexed to this lexically-specific constraint and foots the final syllable fatally violates this constraint and is thus eliminated. In this example, stress must fall on the initial syllable of the stem, as only footed syllables can bear stress. This stress assignment is one syllable to the right of what is predicted by the Regular pattern. At the word-level in (9), as with Regular stress assignment, primary stress is maintained on the same syllable, as candidate (a) does.

---

to NONFINALITY$_1$ because this constraint examines the last syllable of the stem, and this suffix overlaps with that window.

3.3.1 *Class 3 exceptions as Class 1 exceptions*

This analysis also allows for Class 3 exceptions to be treated the same as Class 1 exceptions, shown in (10).

(10) *régimen* "diet" stem-level

/regimen$_1$-/		NON-FINALITY$_1$	FINALSTRESS	NON-FINALITY	TROCHEE
☞	a. *(ré.gi.)men*		*		
	b. *(re.gí.)men*		*		*!
	c. *re.gi.(mén)*	*!		*	

Because this type of word has antepenultimate stress, it is indexed to NONFINALITY$_1$ at the stem-level. Any candidate, like (c), that foots the final syllable fatally violates this constraint. Candidate (b) is also eliminated because it violates TROCHEE, leaving candidate (a) as the optimal stem-level output. As this type of word does not have a TE that is suffixed at the word-level stratum, the fully faithful candidate is optimal at the word-level (not shown).

Class 3 words have not been accounted for in previous analyses, but in the present analysis these exceptional words behave identically to Class 1 exceptions. This reduces the number of exceptional classes and accounts for a set of words that have not been predictably accounted for previously.

3.3.2 *Exceptional words longer than three syllables*

Given the current constraints, exceptional words indexed to NONFINALITY$_1$ that are longer than three syllables are predicted to have antepenultimate stress on the word-stem, yielding preantepenultimate word-level stress. For example, a word like *fotógrafo* "photographer" is predicted to be *fótografo* because preantepenultimate stress would be assigned at the stem-level and then maintained at the word-level due to the high ranking of HEADMATCH. However, this output is ungrammatical because it violates the inviolable three-syllable window. To enforce the three-syllable window at the word-stratum, the following word-level constraint is necessary:

> *EXTENDEDLAPSE: There can be no sequences of more than two adjacent unstressed syllables. Assign one violation for every sequence of more than two unstressed syllables (Gordon 2002).[6]

This constraint allows pairs of adjacent syllables to be stress-less but disprefers sequences of three or more syllables that are all unstressed. The tableaux in (11)–(12) demonstrate how this constraint enforces the three-syllable window in a word like *fotógrafo* "photographer".

---

6. This definition differs from Gordon's (2002) because it is categorical and not gradient.

The stem level tableau in (11) does not show anything novel, as it behaves the same as the other Class 1 words. Candidate (c) is optimal because, as a stem that is indexed to NONFINALITY$_1$, it does not foot the final syllable, and the stressed foot is trochaic.

(11)   *fotógrafo* "photographer" stem-level

/fotograf$_1$-/	NON-FINALITY$_1$	FINALSTRESS	NONFINALITY	TROCHEE
a. *fo.to.(gráf)*	*!		*	
b. *(fo.tó).graf*		*		*!
☞ c. *(fó.to).graf*		*		

Unlike the shorter exceptional words, the faithful candidate is not optimal at the word-level. The faithful candidate (a) in (12) violates *EXTLAPSE because there are three sequential syllables that are all unstressed. *EXTLAPSE must outrank the stress faithfulness constraint, HEADMATCH, to get this correct prediction. Stress must shift to the second or third syllables (*to* or *gra*) to satisfy *EXTLAPSE, but ultimately candidate (c) is optimal because it does not violate the other lower-ranked word-level constraints that (b) and (d) do.

(12)   *fotógrafo* "photographer" word-level

(fó.to).graf + o	*EXTLAPSE	HEADMATCH	NON-FINALITY	TROCHEE
a. *(fó.to).gra.fo*	*!			
b. *(fo.tó).gra.fo*		*		*!
☞ c. *fo.(tó.gra.)fo*		*		
d. *fo.to.(grá.fo)*		*	*!	

Ranking *EXTLAPSE above HEADMATCH at the word-level successfully enforces the three-syllable window that all Spanish words abide by. The inclusion of this constraint does not affect the predictions made for words that are shorter than four syllables, as stress on any syllable of a three syllable or shorter word will satisfy this markedness constraint.

4.   Exceptionality distinguished from ungramaticality

Past approaches of Spanish stress assignment have failed to differentiate between exceptionality and ungrammaticality. For example, Baković's (2016) analysis successfully predicts the exceptional stress of Class 1 words but does so at the expense of also predicting ungrammatical outputs. As mentioned, Spanish words can never have antepenultimate stress when there is a heavy penult. However, in an analysis

that does not incorporate syllable weight into stress assignment in any way (Roca 1988; Baković 2016; Doner 2017, among others), if a nonce input like *rapindo* is marked as exceptional, its optimal output would be *rápindo*, which violates this important generalization.

In contrast, the present analysis does rely on syllable weight for stress assignment in some circumstances, as stress is generally assigned to bimoraic feet at the stem-level. Another circumstance where moraic weight is important for stress assignment in Spanish is for words with heavy penults (i.e., the stem-final syllable is trimoraic or superheavy). Taking this syllable weight into account is necessary to ensure that ungrammatical predictions are not made. Thus, correct stress assignment is predicted for a nonce input like *rapindo*, regardless of whether or not it is marked as exceptional, with the addition of one last markedness constraint that promotes stress on super heavy syllables and is ranked above NONFINALITY$_1$, defined below.

> WEIGHTTOSTRESS$_{\mu\mu\mu}$(WSP$_{\mu\mu\mu}$): Superheavy syllables are stressed. Assign one violation for every superheavy syllable that is unstressed.

Including this general markedness constraint in the stem-level ranking does not affect the predictions made for other words. Within Spanish, not all heavy syllables are stressed, as words like *castigo* "punishment" have an unstressed heavy syllable. However, all superheavy syllables do attract primary stress. For example, words like *comienzo* "start (n.)" always stress the superheavy syllable, as the diphthong is bimoraic and there is a weight-bearing coda.[7] This additional data supports the generalization that words with heavy penults, which at the stem-level will have superheavy stem-final syllables, can never have exceptional stress to the left of this superheavy syllable. The WSP$_{\mu\mu\mu}$ constraint ensures that this inviolable generalization is upheld regardless of whether or not the stems are marked as exceptional. This is shown in the tableaux in (13)–(14) with a nonce word *rapindo$_1$*, which is indexed to NONFINALITY$_1$.

(13)   *rapindo* stem-level

/rapind$_1$-/	WSP$_{\mu\mu\mu}$	NON-FINALITY$_1$	FINALSTRESS	FT-BINμ	NON-FINALITY
a. *(rá.)pind*	*!		*	*	
☞ b. *ra.(pínd)*		*		*	*

---

7.   I assume that the first segment of the diphthong is part of the nucleus rather than an onglide, but refer to Harris (1985) and Martínez-Paricio (2013), among others, for further discussion of this topic.

(14) *rapindo* word-level

ra.(pínd)+ o	HEADMATCH	FT-BINμ	FINALSTRESS
a. (rá.)pin.do	*!	*!	*
☞ b. ra.(pín.)do			*
c. ra.(pín.do)		*!	*

In (13), $WSP_{\mu\mu\mu}$ is fatally violated by candidate (a), as the superheavy syllable does not bear primary stress, thus demonstrating that this constraint must outrank the lexically-indexed version of NONFINALITY₁. Therefore, even though candidate (b) violates NONFINALITY₁ that it is subject to, it is the optimal stem-level output because it stresses the superheavy syllable. Candidate (b) is then the input to the word-level stratum in (14), where HEADMATCH pressures stress to remain on the same syllable as in the input, making candidate (b) the ultimate winner. Hasse diagrams of the final constraint rankings for each stratum are shown in (15)–(16), respectively.

(15) Stem-level constraint ranking

```
              WBP
               |
             WSPμμμ
               |
          NONFINALITY₁
               |
           FINALSTRESS
               |
            FT-BINμ
            /       \
   NONFINALITY      TROCHEE
```

(16) Word-level constraint ranking

```
        *EXTENDEDLAPSE
               |
    HEADMATCH      FT-BIN
         \    X    /
   NONFINALITY    TROCHEE
                     |
                FINALSTRESS
```
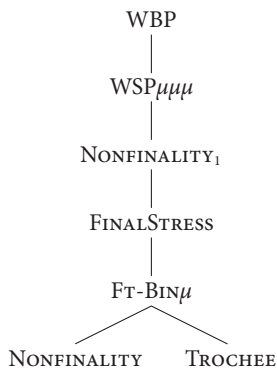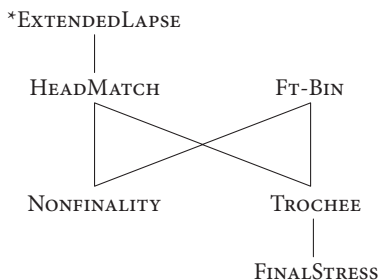
This analysis differs from previous accounts of Spanish stress in that it allows for exceptionality in stress assignment but does not predict the possibility of an ungrammatical output. This is in contrast specifically to Baković's (2016) analysis. His

analysis may seem simpler in that it utilizes a standard parallel OT framework and requires fewer constraints, but this comes at the cost of predicting ungrammaticality. This is demonstrated below.

Prior to explaining the incorrect prediction made by Baković's (2016) analysis, I introduce his constraints and how they interact with Regular and Class 1 words.

FINALSTRESS: Stress is stem final; assign one violation for every instance that the stressed syllable precedes the unstressed syllable within the stem.
NONFINALITY: Stress is not final in the stem (1 violation) or word (2 violations).
NONFINALITY$_1$: Stress is not final in the stem (1 violation) or word (2 violations) for lexical items that are indexed to this constraint.

As in the present analysis, he models the Regular stress pattern by ranking FINALSTRESS >> NONFINALITY; this is shown in (17). Also similar to the present analysis, Class 1 exceptions are accounted for with a lexically-indexed version of his NONFINALITY, NONFINALITY$_1$, and the ranking NONFINALITY$_1$ >> FINALSTRESS >> NONFINALITY. This is shown in (18).

(17)   Baković's (2016) analysis of a Regular word

/herman-a/	FINALSTRESS	NONFINALITY
☞ a. *hermán-a*		*
b. *herman-á*		**!
c. *hérman-a*	*!	

(18)   Baković's (2016) analysis of a Class 1 exception

/rapid-o/1	NONFINALITY$_1$	FINALSTRESS	NONFINALITY
☞ a. *rápid-o*		*	
b. *rapíd-o*	*!		*
c. *rapid-ó*	**!		**

In (17), FINALSTRESS is satisfied by both candidates (a) and (b) because stress does not appear to the left of the stem-final syllables, as in (c). Candidate (a) is optimal because Baković's (2016) version of NONFINALITY penalizes word-final stress more than stem-final stress. In (18), as NONFINALITY$_1$ is fatally violated by candidates (b) and (c), candidate (a) is the correct winner.

Although these constraints and their respective ranking correctly predict stress for many Spanish words, Baković's (2016) analysis, and the many others that do not take syllable weight into account, conflates exceptionality with ungrammaticality and makes incorrect and unattested predictions for potential words with heavy penults.

In a constraint-based framework like OT, Richness of the Base (ROTB) must be assumed, which states that inputs cannot be restricted. Thus, any input is possible

but the constraints and their ranking must yield grammatical outputs, regardless of any constraint indexation. Under Baković's (2016) analysis, if an input with a heavy penult is indexed to NONFINALITY$_1$, it will be predicted to have antepenultimate stress, which is ungrammatical. This is shown below in (19). As there is nothing to restrict this input from emerging, it is highly problematic that an ungrammatical output is selected as optimal. This is also problematic because experimental evidence has shown that Spanish speakers, when assigning stress to novel word forms, do not produce antepenultimate stress if a word has a heavy penult (Waltermire 2004).

(19) Incorrect prediction of Baković's (2016) analysis

	/rapind-o/	NONFINALITY$_1$	FINALSTRESS	NONFINALITY
☹	a. *rápind-o		*	
	b. *rapínd-o*	*!	*	
	c. *rapind-ó*	*!*		**

This result is in contrast to the results of the present analysis, which do not select the incorrect output for a word form of this shape (i.e., *rápindo) because the analysis is constrained to predict the correct output that abides by the two Spanish inviolable stress generalizations. It is impossible to make this prediction in a system that does not take syllable weight into account in some manner. As I have shown, it is clearly possible to rely on syllable weight for stress assignment under some contexts, while still making the correct predictions for words that appear to ignore the pressures of quantity sensitivity (e.g., *castigo* "punishment"), and while *not* predicting incorrect forms for other possible inputs (e.g., *rapindo*).

4.1   Truncations

Additional support for the present analysis comes from stress in truncations. In Spanish, words are truncated in a systematic way. A word like *profesor* "professor" can be shortened to *profe* "prof". The stem of a truncated word is the full word-stem, meaning that the stem of *profe* is *profesor-*. The tableaux of each stratum for this word are shown in (20)–(21).

(20) *profe* "prof" stem-level

	/profesor-/	FINALSTRESS	FT-BINμ	TROCHEE	NONFINALITY
	a. *(pró.fe.)sor*	*			
	b. *(pro.fé)sor*	*		*!	
☞	c. *pro.fe.(sór)*				*
	d. *pro.(fe.sór)*		*!		*

(21) *profe* "prof" word-level

(pro.fe)	HEADMATCH	FT-BINµ	TROCHEE	NONFINALITY	FINALSTRESS
a. *(pró.)fe*		*!			*
☞ b. *(pró.fe)*				*	*
c. *(pro.fé)*			*!	*!	

The stem-level stratum assigns Regular stem-final stress. The truncation then happens at the word-level. Because the stressed syllable of the stem is no longer part of the word-level stratum, there is no stress to be faithful to. Thus, none of the candidates violate HEADMATCH. Candidate (b) is the winner because it satisfies both FT-BINµ and TROCHEE, which candidates (a) and (c) violate.

This differs from the analysis of truncations put forth by Baković (2016). His analysis states that truncations, which are formed from Regular class words and all behave systematically, are all exceptional and indexed as such. As syllable weight and foot structure do not play a role in stress assignment under his approach, these Regular truncated words are predicted to have final stress unless marked as exceptional. The present Stratal OT analysis, however, correctly predicts antepenultimate stress while treating truncations as Regular (i.e., they are not indexed).

5. Remaining words

There is a small set of words that the analysis above cannot account for, such as *imágen* "image" and *resúmen* "summary", which maintain their original stress from Latin. The present analysis as it stands would assign antepenultimate stress. However, these could be accounted for by making them into a subset of the Class 1 exceptions, which would mean that they are subject to NONFLINATLY$_1$. Only this subset of words that would bear a second index would be subject to a lexically-indexed version of IAMB, ranked above the general TROCHEE at the stem level. The alternative to this would be to argue that these few words have lexical stress. Importantly, neither of these options affect the predictions made for other words given the above analysis.

6. General discussion and conclusions

This paper has presented a novel analysis of Spanish stress assignment, which utilizes lexically-indexed constraints in a Stratal OT framework. Unlike all other previous analyses of these data, the present analysis is able to successfully account for exceptionality in Spanish stress assignment without predicting ungrammatical

forms that violate the two exceptionless Spanish stress generalizations. This is due to the unique proposal that Spanish stress assignment is sensitive to the weight of the morphologically-derived word-stem when a superheavy syllable is present. Additionally, the present analysis is able to account for more of the Spanish lexicon, including Class 3 exceptions, without additional machinery; no previous analysis has attempted to account for this set of words, yet the present analysis collapses these words with Class 1 exceptions. The predictions made by the proposed grammar could not be made in an analysis that does not value syllable weight, and thus, the current proposal supports some degree of quantity sensitivity in Spanish at the stem-level, in order to differentiate between exceptional and ungrammatical stress assignment in Spanish.

Acknowledgements

References

Baković, Eric. 2016. Exceptionality in Spanish stress. *Catalan Journal of Linguistics* 15. 9–25. https://doi.org/10.5565/rev/catjl.182

Bermúdez-Otero, Ricardo. 2008. *Stratal Optimality Theory*. Oxford: Oxford University Press.

Doner, Julianne. 2017. Spanish stress and lexical accent across syntactic categories. *Probus* 29(2). 233–285. https://doi.org/10.1515/probus-2017-0001

Dunlap, Elaine R. 1992. Issues in the moraic structure of Spanish. Amherst MA: University of Massachusetts Amherst dissertation.

Eisner, Jason. 1997. What constraints should OT allow. Paper presented at the 71st Annual Meeting of the Linguistic Society of America, Chicago, IL, January 2–5. *Available as ROA-204 from the Rutgers Optimality Archive*. Handout.

Gordon, Matthew. 2002. A factorial typology of quantity-insensitive stress. *Natural Language & Linguistic Theory* 20(3). 491–552. https://doi.org/10.1023/A:1015810531699

Harris, James W. 1982. *Syllable structure and stress in Spanish. A nonlinear analysis*. Cambridge, MA: MIT Press.

Harris, James W. 1985. Spanish diphthongisation and stress: A paradox resolved. *Phonology* 2(1). 31–45. https://doi.org/10.1017/S0952675700000373

Harris, James. W. 1989. Sonority and syllabification in Spanish. *Studies in Romance Linguistics*, 139–153. https://doi.org/10.1075/cilt.60.10har

Hualde, José I. & Marianna Nadeu. 2014. Rhetorical stress in Spanish. In Harry van der Hulst (ed.), *Word stress: Theoretical and typological issues*, 228–252. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139600408.010

Jarosz, Gaja. 2018. Indexed morphemes and locality conditions of Polish yer deletion. Paper presented at the 26th Manchester Phonology Meeting, Manchester, UK, May 24–26.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17(2–4). 351–366. https://doi.org/10.1515/tlir.2000.17.2-4.351

Lipski, John M. 1997. Spanish word stress: The interaction of moras and minimality. In Fernando Martínez-Gil & Alfonso Morales-Front (eds.), *Issues in the phonology and morphology of the major Iberian languages*, 559–593. Washington, DC: Georgetown University Press.

Martínez-Paricio, Violeta. 2013. The intricate connection between diphthongs and stress in Spanish. *Nordlyd* 40(1). 166–195. https://doi.org/10.7557/12.2505

McCarthy, John J. 2000. The prosody of phase in Rotuman. *Natural Language & Linguistic Theory* 18(1). 147–197. https://doi.org/10.1023/A:1006342918830

McCarthy, John J. 2003. OT constraints are categorical. *Phonology* 20(1). 75–138. https://doi.org/10.1017/S0952675703004470

McCarthy, John J. & Alan S. Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language & Linguistic Theory* 8(2). 209–283. https://doi.org/10.1007/BF00208524

Pater, Joe. 2000. Non-uniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17(2). 237–274. https://doi.org/10.1017/S0952675700003900

Pater, Joe. 2009. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In Steve Parker (ed.), *Phonological argumentation: Essays on evidence and motivation*, 123–154. Sheffield, UK, Equinox.

Prince, Alan & Paul Smolensky. 1993/2004. *Optimality Theory: Constraint interaction in Generative Grammar*. Malden, MA: Wiley-Blackwell.

Roca, Iggy. 1988. Theoretical implications of Spanish word stress. *Linguistic Inquiry* 19(3). 393–423.

Waltermire, Mark. 2004. The effect of syllable weight on the determination of spoken stress in Spanish. In Timothy Face (ed.), *Laboratory Approaches to Spanish Phonology*, 171–191. New York: Mouten de Gruyter.

Index

This volume brings together selected papers from the 48th annual
Linguistics Symposium on Romance Languages, held at York
University in Toronto, Canada, in April 2018. It presents original
research on a wide variety of Romance languages both past (Latin,
Old Catalan, Old Iberian Romance, Old Spanish, Old Portuguese,
and West-Iberian Medieval Latin) and present (Brazilian Portuguese,
Catalan, French, Picard, Portuguese, Romanian, and Spanish) along
with a number of contemporary dialects, including Basque Country
Spanish, Dominican Spanish, Maine French, Neapolitan, and Picardie
French. Divided into four sections – Interfaces, Bridging issues at
the CP-TP-vP levels, Bridging issues at the PP-DP levels, and Bridging
issues in linguistics – the volume gives researchers and advanced
students access to contemporary issues and novel ideas bridging
across various areas of Romance linguistics (e.g., morphology, syntax,
semantics, phonology, sociolinguistics, first and second language
acquisition).

ISBN 978 90 272 1084 5

9 789027 210845

JOHN BENJAMINS PUBLISHING COMPANY