Romance Languages and Linguistic Theory

16

Romance Languages and Linguistic Theory 16

Selected papers from the 47th Linguistic Symposium on Romance Languages (LSRL), Newark, Delaware

edited by
Irene Vogel

John Benjamins Publishing Company

Romance Languages and Linguistic Theory 16

Romance Languages and Linguistic Theory (RLLT)

ISSN 1574-552X

The yearly events of both 'Going Romance' and 'Linguistic Symposium on Romance languages' feature research in formal linguistics of Romance languages, in the domains of syntax, morphology, phonology and semantics. Each volume brings together a peer-reviewed selection of papers that were presented at one of the meetings, aiming to provide a representation of the spread of topics at that conference, and of the variety of research carried out nowadays on Romance languages within theoretical linguistics.

For an overview of all books published in this series, please see benjamins.com/catalog/rllt

Editor

Frank Drijkoningen
Utrecht University

Volume 16

Romance Languages and Linguistic Theory 16. Selected papers from the 47th Linguistic Symposium on Romance Languages (LSRL), Newark, Delaware
Edited by Irene Vogel

Romance Languages and Linguistic Theory 16

Selected papers from the 47th Linguistic Symposium on Romance Languages (LSRL), Newark, Delaware

Edited by

Irene Vogel

John Benjamins Publishing Company

Amsterdam / Philadelphia

DOI 10.1075/rllt.16
Cataloging-in-Publication Data available from Library of Congress:
LCCN 2020007370 (PRINT) / 2020007371 (E-BOOK)

ISBN 978 90 272 0555 1 (HB)
ISBN 978 90 272 6118 2 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of Contents

Introduction

Irene Vogel
University of Delaware

Since its inception in 1971, the Linguistic Symposium on Romance Languages (LSRL) has played a central role in the field of Romance linguistics, and indeed, linguistics in general. Beginning shortly after the introduction of generative linguistics, LSRL has always been on the cutting edge of linguistic theory, drawing its subject matter and data from the family of Romance languages. While it may seem that this is a rather restrictive area of investigation, generally involving well-studied languages, it has also offered opportunities for the advancement of our understanding of aspects of language and linguistics that would not be available with more distantly, or unrelated, languages.

Obvious cases involve issues related to historical development and comparative studies within all of the sub-disciplines of linguistics. With the more recent inclusion of studies of smaller dialects, the Romance linguistic terrain now also offers the possibility of cutting-edge investigation of "micro-variation" or "micro-comparisons". In addition, given earlier colonial history involving the spread of Spanish, French and Portuguese to many areas beyond the original Romance region, it is now possible to examine how these languages have diverged, and indeed continue to diverge. We are also able to investigate how the same languages have been influenced by a range of different contact languages, past and present, including the languages of African slaves, indigenous languages of the Americas, English. Methodological and technical advances in recent years have also introduced different perspectives and approaches to our study and understanding of the Romance languages, including socio-phonetics, computational, psycholinguistic and biological (e.g., DNA) analyses.

The chapters in this book comprise a selection of sixteen papers and posters presented at the 47th meeting of the Linguistic Symposium on Romance Languages (LSRL 47), held at the University of Delaware, April 20–23, 2017. LSRL 47 adopted the theme of "bridges" – bridging research approaches and directions across languages, departments and disciplines. The goal was to include contributions not only on standard language varieties, but also lesser known, minority and possibly endangered varieties, and to include contributions that not only use the

https://doi.org/10.1075/rllt.16.int

usual descriptive and theoretical approaches of linguistics, but also more interdisciplinary approaches including psycholinguistics, neurolinguistics and computational linguistics. Furthermore, it was felt that the bridging theme was also firmly in line with LSRL's history of presenting high level linguistic research, and embracing cutting-edge developments in the field.

The five plenary speakers at LSRL, Drs. Gregory Guy, Paola Merlo, Thierry Nazzi, Pilar Prieto and Armin Schwegler, were invited precisely to reflect the broad and inclusive range of issues pertaining to Romance languages currently being addressed across different branches of linguistic research. The chapters included in this book are also reflective of the diversity of topics, approaches and languages represented at the Conference. Of the 131 abstracts submitted for the conference, a total of 90 were accepted for either oral or poster presentation, including two special sessions, focusing on Second Language Acquisition of Romance Languages and a Cross-Linguistic Investigation of Romance Clitics. Subsequently, twenty-four papers were submitted for publication based on oral and poster presentations. They were all reviewed anonymously, and sixteen were accepted for publication, including three contributed by plenary speakers. Given the range of topics discussed in the chapters of this book, they are thus presented in alphabetical order by author, since they do not all fall into several neat categories.

In the first Chapter, "Variability in French Word-Final Schwa: The Effect of Focus", Angeliki Athansopoulous and Irene Vogel investigate prosodic effects on schwa manifestation in French. In this experimental study, final schwas were examined in particularly favorable contexts, following a consonant cluster violating the Sonority Sequencing Principle, and before another word beginning with a consonant (e.g., […*lisiblə#set*…] '…readable this…'). Since focus often induces "hyper-articulation", the authors compared the occurrence of schwa in focus and non-focus contexts, the former predicted to exhibit more schwa production than the latter. It was indeed found that the stronger boundary associated with focus resulted in increased schwa production in the target constructions. Additionally, it was observed that within this overall pattern, there were also some differences in the rates of schwa usage among the individual participants.

The second Chapter, "Agreement and Pronouns: Implications for Partial Control", by J.-Marc Authier and Lisa A. R. Reed, uses the two-step theory of Agree to compare the workings of agreement mismatches involving French 'on' interpreted as first person plural with those involving PRO in sentences exhibiting partial control. Authier and Reed show that, unlike first person plural 'on', partially controlled PRO leaves no morphological or semantic agreement footprint. This leads them to conclude that whatever induces the inference of plurality associated with partially controlled PRO likely arises through the sort of process of pragmatic associative reference advocated.

In the third Chapter, entitled "Person/Number exponents in imperative-enclitic contexts," Teresa Cabré and María Ohannesian analyze the systematic deletion of the Person/Number exponent in imperative-reflexive enclitic contexts in Spanish and Barcelona Catalan. They show that when identical values for Person/Number features collide, one element must be eliminated. Although Spanish deletes the right edge element of the verb, and Catalan deletes the left edge element of the clitic, the authors propose that both options arise from the same principles. They argue in favor of an OT analysis since the phonological deletions are morphologically conditioned, and OT allows them to combine both the phonological and morphological restrictions, and ultimately account for the different patterns with the same set of constraints.

The fourth Chapter, by Isabelle Charnavel, entitled "French causal *puisque*-clauses in the light of (not)-at-issueness", examines the discourse status of French causal constructions introduced by *puisque* 'since'. *Puisque*-clauses are associated with two implications: the relation expressed by *puisque* and the content of the clause. Several diagnostics show that neither implication is at-issue and that the two implications belong to different types of projective content based on both syntactic and lexical considerations. That is, the relation expressed by *puisque* is not at-issue because *puisque*-clauses modify high (evidential and speech act) phrases that are not at-issue; and the content of the *puisque*-clause is not at-issue because *puisque* is lexically factive, (i.e., selects true facts). It is thus proposed that *puisque* exemplifies an unrecognized type of double presuppositional trigger.

In the fifth Chapter, "Geminates and vowel laxing in Quebec French", by Anaële Cuerrier & Charles Reiss, it is demonstrated that Quebec French has geminates that close preceding syllables and thus create the conditions for laxing of high vowels, as in other closed syllables. Since the existence of such geminates was not widely recognized, the presence of the lax vowels in question had previously been attributed to an otherwise unattested form of obligatorily long-distance vowel harmony.

In Chapter 6, "Number as an adjunct in Romance", Sonia Cyrino and M. Teresa Espinal, diverge from previous claims, and argue that Spanish (and other Romance languages such as Catalan and some Lunigiana dialects) patterns with languages such as Brazilian Portuguese and French. Specifically, they propose that plural marking is a syntactic adjunct to D (i.e., a categorized *d* root) by default, and that (plural) Number on other constituents within the nominal domain must be considered solely as the output of morphophonological agreement or concord.

Chapter 7, "A sociophonetic investigation of Mexico City Spanish vowel reduction," by Meghan Dabkowski, presents an instrumental acoustic analysis of vowel reduction in the variety of Spanish spoken in Mexico City, and examines linguistic and social factors conditioning its occurrence. The analysis revealed

that speakers use two complementary strategies for reducing vowels: shortening and voice weakening. The findings, moreover, corroborate earlier claims based on impressionistic explorations that social factors do not contribute to the occurrence of either shortening or voice weakening.

In Chapter 8, "Computational Quantitative Syntax: The Case of Universal 18", Kristina Gulordava and plenary speaker Paola Merlo present quantitative evidence regarding Greenberg's word order universal, Universal 18. They show that corpus data confirms the dispreference described by this universal for the word order combination where adjectives precede nouns but numerals follow them (i.e., Adj-N but N-Num). Based on corpus counts, Gulordava and Merlo argue that this dispreference is better explained as a constraint expressed at the level of individual structures rather than at the level of the dominant orders. Finally, they discuss how to integrate this bias against Adj-N-Num orders within the grammar.

Chapter 9, by plenary speaker Gregory Guy, "Trajectories of change in Spanish and Portuguese in the Americas" traces the directions, causes, and outcomes of change in Spanish and Portuguese in the Americas across the five centuries since Columbus and Cabral. Guy notes that in addition to spontaneous innovations, such as the emergence of "zheismo" in Argentine Spanish and the first person plural pronoun *a gente* in Brazilian Portuguese, there have been significant changes motivated by language contact. For example, early contact with African languages under the practice of slavery is a likely source of numerous shared characteristics of popular varieties of Caribbean Spanish and Brazilian Portuguese, including coda consonant deletions, erosion of agreement marking, and elevated levels of overt subject pronouns. In the contemporary world, we now observe dialect leveling, expanded education, and contact with English driving other changes.

In Chapter 10, "Clause Typing and Quebec French *-tu*", Daniela Isac and Justin Royer propose that the distributional properties of *-tu*, which spans three types of clauses – interrogatives, exclamatives and declaratives indicating surprise – can be explained if one assumes that *-tu* is the overt realization of two features of the contexts share: a Polarity and a Modality feature. Specifically, they propose that the Polarity feature shared by the three types of clauses, and instantiated by *-tu*, is valued as neither negative nor positive, while the type of Modality that cuts across the contexts that license *-tu* is epistemic. Thus *-tu* is licensed only in contexts in which the epistemic commitments of both the Addressee and the Speaker are encoded, and the epistemic commitment of the Speaker is lower than that of the Addressee (i.e., the Speaker is more ignorant than the Addressee).

In Chapter 11, "The syntax of superlative phrases in Romance", Nicoletta Loccioni argues that the adnominal position of superlative forms in Romance is not the result of NP movement (or lack thereof) crossing the adjective, and thus differs from that of ordinary adjectives. Moreover, prenominal and postnominal

superlatives involve different derivations. It is proposed that the former (and arguably elliptical superlatives) are derived through movement of the adjective to a scope position in the left periphery of the DP, where they cannot be by-passed by the NP. The latter are generated as reduced-relative-clause predicates, which necessarily extrapose in Romance. They are analyzed as maximalizing relative clauses over degrees, and are equipped with a max operator over degrees which is spelled out as an overt definite determiner in French but left unpronounced in languages like Italian and Spanish.

Chapter 12, "On recomplementation, high adverbs and V-movement in Spanish", by Gabriel Martínez Vera, establishes a contrast between Latin American Spanish (LAS) and European Spanish (ES) with regard to double-complementizer constructions. Specifically, Martínez Vera demonstrates that while there is an overt low complementizer *que* in ES, there is a null low complementizer in LAS. Moreover, constructions with the latter in LAS contrast with a similar construction in ES where low *que* is absent, and it is proposed that this contrast is tied to a difference in the height of V-movement and adverb placement in both LAS and ES.

In Chapter 13, "Articles in an article-less language: Romanian-Serbian code-switching", Vanessa Petroj investigates patterns of Romanian-Serbian code-switching. Specifically, she examines the interactions between Romanian definite articles, Romanian and Serbian adjectives, and Romanian and Serbian nouns. Although code-switching between the two languages in question often results in language mixing throughout the entire structure, Petroj's focus here is on the combination of the absence of articles in Serbian, and the affixal nature of the Romanian definite article, a property that is atypical among the Romance languages.

In Chapter 14, "Split auxiliary selection with affected subjects in Old Majorcan Catalan," Sebastià Salvà i Puig addresses the issue of split auxiliary selection (both person-driven and event-driven) in Old Majorcan Catalan. He specifically investigates constructions where, instead of *haver* 'have', *esser* 'be' is used as the auxiliary verb for compound tenses in first and second person forms, not only with unaccusative verbs, but also with transitive and unergative verbs (e.g., *La som vista* 'I have seen her'). It is proposed that the constructions in question are "unaccusative" in a broad sense, with affected subjects derived from a lower position (i.e., at least [Spec, ProcessP], where the argument is interpreted as an undergoer).

Chapter 15, by plenary speaker Armin Schwegler, is entitled "Palenque (Colombia): African language survivals and their identification". In this chapter, Schwegler documents the revitalization of African language survivals in the Afromaroon community of Palenque (Colombia), where creole and Spanish have been spoken side by side for almost four centuries. Prior to the close of the 20th century, lexical and other Africanisms in Palenque were surprisingly scant in everyday parlance. More recently, current sentiments of *negritud* ('Black awareness') and other

factors have, however, given archaic *lexicalia "africana"* a new lifeline, thereby triggering the rebirth of words that are now used as a genuine index of Palenqueros' maroon and Afro-Colombian identity.

In the final chapter, Chapter 16, "Inflected infinitives in Galician", Michelle Sheehan, Jeffrey Blokzijl, M. Carmen Parafita Couto use an audio-questionnaire, supplemented by spoken corpus data, to investigate the acceptability of Galician inflected infinitives in different syntactic contexts. Despite normative pressures against the use of inflection in the case of exhaustive local subject control, examples are attested in corpora and accepted by some speakers in the audio-questionnaire. Uncontrolled inflected infinitives are widely rejected by Galician speakers, whereas partial control examples are accepted by many more. Overall, the results reveal that inflected infinitives in Galician are acceptable in a broader range of contexts than previously reported, and specifically, in a different range of contexts than observed in closely related Portuguese.

As this brief summary of the chapters indicates, this volume comprises not only descriptive and theoretical studies of languages, but also experimental studies addressing issues in syntax, phonetics-phonology and sociolinguistics. The languages investigated include not only European varieties of French, Spanish and Portuguese, but also Quebec French, several varieties Spanish spoken in the Americas and the Caribbean, and Brazilian Portuguese. Less widely spoken languages including Romanian, Galician and Catalan are also represented among the contributions, from both synchronic and diachronic perspectives. Romance languages in contact with other languages and bilingualism are now also integral aspects of the field, as reflected in this volume as well, including less well-known cases of contemporary contact of Serbian with Romanian, and earlier contact of African languages with Spanish and Portuguese. Readers interested in Romance languages and linguistics, as well as general linguistics, sociolinguistics, experimental linguistics, and innovative approaches to linguistic investigation will all find chapters of interest in this book.

The colleagues who read the abstracts for the conference and the papers for this volume are too numerous to thank individually, but their help was invaluable. The LSRL 47 conference itself would not have been possible without the generous contributions from NSF (grant #1644342) and several University of Delaware bodies: the Department of Linguistics and Cognitive Science, the College of Arts and Sciences and the Institute for Global Studies.

Variability in French word-final schwa

Effects of focus and speaker

Angeliki Athanasopoulou & Irene Vogel
University of Calgary / University of Delaware

The production of schwa in French is highly variable. Some of this variability is predictable based on sociolinguistic (e.g., dialects), phonological (e.g., segments, word position), and stylistic (e.g., reading) factors. We investigate the effect of prosodic structure on the production of word-final schwa in Parisian French, while controlling for the other factors. We also investigate variability in schwa production within and between speakers of the same variety. Our findings show that prosody does affect the production of schwa: the stronger the prosodic boundary, the more probable schwa production is. While this is true for all the speakers, we also found considerable variation in the rates of schwa production, despite our controlling for dialect and style.

Keywords: Parisian French, word-final schwa, prosodic boundary, focus, variability

1. Introduction

As is well known, a variety of factors influence the production of schwa in French, including phonological, morphological, lexical, sociolinguistic, and stylistic considerations. As such, the topic of French schwa has generated a number of competing analyses that attempt to explain when a schwa is produced. Since different studies use different designs and focus on different factors (e.g., phonological, sociolinguistic), it is difficult to compare the results and assess the effects of each factor independently of the others. Despite such differences, it is recognized across studies that there is considerable variability in whether or not schwa is produced, also due simply to speaker differences (e.g., Bayles, Kaplan, & Kaplan 2016).

In the present study, we focus on the Parisian dialect of French, and investigate the production of word-final schwa in a context that is especially favorable for its production. Our main goal is to isolate the prosodic effects from other factors that contribute to the variability of schwa production, and thus specifically assess

https://doi.org/10.1075/rllt.16.01ath

the role of prosody in this variability. We examine the data further, considering the variability of schwa production in individual speakers.

In Section 2, we provide a brief overview of the main issues that contribute to the complexity of schwa production patterns. Sections 3 and 4 present our hypotheses and the methodology we used in the present investigation, respectively. The results are presented in Section 5. Section 6 offers a general discussion, followed by conclusions in Section 7.

2. French schwa

It is well known that dialect plays a crucial role in schwa production (e.g., Gess, Lyche, & Meisenburg 2012), with Southern varieties of French exhibiting schwa more frequently than the Northern (Parisian) variety examined here. For example, while the word *facile* 'easy', is produced with a schwa (*facil*[ə]) in the South, it is pronounced without schwa (*facil*[-]) in the North (Eychenne 2015). Other sociolinguistic considerations such as the age, sex, and social class of the speakers also affect schwa production (Léon 1993; Hansen 1994, 1997; Carton 1999), as does speech style. The more formal styles (e.g., reading) present more frequent schwa production than less formal ones (Lucci 1983; Hansen 1994, 2012; Eychenne 2006, 2009; Pustka 2009; Hansen & Hansen 2003), as well as much lower variability (Bayles et al. 2016).

Certain phonological factors also affect schwa production, such as the position of the schwa in a word (Côté 2001; Hansen 1994, 2012; Walter 1990; Fónagy, 1989), an adjacent prosodic boundary (Côté 2001, 2007; Lacheret & Lyche 2008; Eychenne 2005; Hansen 1997, 2012), and the segmental or syllabic context around the schwa (Côté 2001; Dell 1985; Tranel 2000; Pulgram 1961; Hansen 2012; Bayles et al. 2016). The last factor (segmental context) has drawn considerable attention in the literature, in particular the well-known *loi des trois consonnes* (LTC). According to the LTC, the schwa may be produced after two consonants, but not after one, when followed by another consonant. So, in the phrase *aigl*[ə] *domestique* 'tame eagle', the schwa is produced because it is preceded by two consonants [gl] and followed by a third one, but in *mouch*[-] *domestique* 'tame fly', the schwa is not produced because it is preceded by a single consonant, [ʃ], and also followed by a single consonant. Similarly, the schwa is not produced before a vowel, as in *aigl*[-] *orange* 'orange eagle'.

In some analyses, the context effects are said to be "purely sequential" (Côté 2001, p. 88), the result of the specific sequence of segments surrounding the schwa (e.g., Dell 1985), while in other analyses, the effects are considered "explicitly syllabic" (Côté 2001, p. 87), reflecting the syllabification of the segments (e.g.,

Pulgram 1961; Tranel 2000). It has been noted that even in items like *aigle* 'eagle', there is a good deal of variability among speakers with regard to schwa production, and thus, the syllabic account cannot explain all of the observed variability in the data (Côté 2001). While the LTC captures a substantial portion of the variability in schwa production, Côté (2001) shows that it, too, falls short since it is too general, and additional conditions are needed to capture the full range of variability. For example, the schwa may be omitted in the LTC context when the second consonant is less sonorous than the adjacent consonants, when the second consonant is a stop and its perceptual cues are not obscured by the adjacent consonants, when the schwa is at a large prosodic boundary, and when the schwa is in a clitic. Bayles et al. (2016) show that while Côté's additional conditions better capture the variability in schwa production than the LTC on its own, there is still, as a minimum, speaker variability that must be accounted for.

In addition to the segments immediately surrounding the schwa, the position of the schwa in the word affects its production, but only in non-LTC contexts (e.g., VC_C). That is, schwa is more likely to be produced in word-initial syllables (e.g., *f*[ə]*nêtre* 'window'), than word-medial syllables (e.g., *sam*[ə]*di* 'Saturday') and word-final syllables (e.g., *(je) laiss[ə]* 'I leave'), where, instead, it is almost categorically dropped (Hansen 2012; Bayles et al. 2016). By contrast, in LTC contexts (e.g., CC_C), schwa production does not substantially differ across word positions. In all three, schwa is produced the majority of the time: 67%–68% in conversation, and even more, 84%-100%, in reading (Hansen 2012). Bayles et al. (2016) show a similar trend, but with somewhat different degrees of schwa production: in conversation, schwa production ranged from 30% to 100% for individual speakers, but in reading, no variability was observed, with all speakers producing the schwa at 100%.

The last phonological factor we must consider is the prosodic structure. The initial formulation of the LTC does not distinguish between different prosodic boundaries that may be present following the schwa. Côté (2001), however, shows that the distribution of schwa in the CC_C context is not the same when, for example, there is no prosodic boundary, CC_C (e.g., *justement* 'justly') and when there is a Phonological Word (PW) boundary, CC_ $]_{PW}$ C (e.g., *le juste*$]_{PW}$ *ment* 'the just (one) lies'). In the first case, the schwa is almost always produced (*just*[ə]*ment*), but in the second, it is optional (*le just*([ə]) $]_{PW}$ *ment*). Moreover, the effect of prosodic boundaries is cumulative from the lower (e.g., PW), to the higher boundary (e.g., Intonational Phrase, IPh), thus according to Côté (2001), there is also a difference in schwa production between the phrases *infecte manteau* 'stinking coat' and *l'insecte, mets-le là* 'the insect, put it there'. In the former, the schwa is adjacent to a PW boundary, and its production is obligatory (*infect*[ə] $]_{PW}$), while in the latter, the schwa is adjacent to a higher boundary (IPh), and its production is optional

(*l'insect*([ə]) $]_{IPh}$). Côté's (2001) proposal that the stronger the prosodic boundary adjacent to a schwa, the less likely it is for the schwa to be produced, is consistent with Dell's (1977) much earlier finding. (See Côté (2001) for the prosodic analysis of Dell's (1977) design.) The schwa was found to be more likely to be produced at a PW boundary than at a Phonological Phrase (PPh) boundary. Côté's (2001) proposal is in conflict, however, with Hansen's (1997) analysis of the "prepausal schwa," according to which a word-final schwa typically appears before a pause, often at the end of a sentence (whether the schwa is etymologically present or not). For example, while schwa is not normally produced in Northern French at the end of word, it may be produced before a pause (Hansen 1997, 2012; Eychenne 2005): e.g, *maîtris*[-] 'mastery' is produced with a schwa in *J'ai fait mon mémoire de maîtris*[ə]*, sûr euh le système…* 'I did my research paper, on uh the system…' (Hansen 1997, p. 166). Hansen (2012) shows, moreover, that the higher the prosodic boundary is in the prosodic hierarchy, the more likely it is for the schwa to be produced, since in both conversational and reading styles, the schwa is, in fact, produced more frequently before a "pause marquée" (which is possibly an IPh boundary), than a shorter pause (possibly a PPh).

3. Present study

The present study is part of a larger cross-linguistic study of stress and focus (Vogel, Athanasopoulou, & Pincus 2016) where we investigate, among other phenomena, the differences between the typical final stress in French (e.g., *immigrer* [imi'gre] 'to immigrate'), and the less common penultimate stress induced by the presence of a final schwa (e.g., *immigre* [i'migrə] '(he) immigrates'), as well as the effect of focus on their properties. We examine here a subset of the data from the larger corpus, the words that end with schwa (i.e., non-final stress words). Specifically, we investigate to what extent the schwa is produced in the LTC context CC_#C, and whether different prosodic boundaries, introduced with the different focus conditions, affect the rate of schwa production in this segmental context.

3.1 Hypotheses

To test the effects of prosody on schwa production, each word was placed in a focused and a non-focused context. In the focused context, there is a strong prosodic boundary (IPh) after the schwa, CC_ $]_{IPh}$ C, while in the non-focused context, there is a weaker boundary (PPh) after the schwa, CC_ $]_{PPh}$ C. As noted above, there are two competing claims in the previous studies regarding the effect of prosody on schwa: according to Côté (2001), the stronger the prosodic boundary, the less likely it is for the schwa to be produced, but according to Hansen

(2012), the stronger the prosodic boundary, the more likely it is for the schwa to be produced. We thus test two corresponding competing hypotheses, shown in (1).

(1) Hypotheses testing the effect of prosody on schwa production.
   a. Hypothesis 1: There is less schwa production at an IPh boundary than at a PPh boundary. (compatible with Côté (2001))
   b. Hypothesis 2: There is more schwa production at an IPh boundary than at a PPh boundary. (compatible with Hansen (2012))

In addition, even though we controlled for a number of factors to minimize the variability in schwa production, and we selected a segmental context where schwa production is highly favored (based on the LTC and Côté's (2001) conditions), we also took into consideration the finding of Bayles et al. (2016), and we thus additionally considered the possibility that schwa production would still be somewhat variable depending on the speakers. We thus analyzed the data from each speaker separately, to test Hypothesis 3, provided in (2).

(2) Hypothesis on the variability of schwa production.
   a. Hypothesis 3: The production of schwa will be variable both within and between the speakers in each prosodic boundary condition.

4. Method

4.1 Participants

To reduce dialectal and other sociolinguistic variability, the participants were 9 native monolingual speakers of Parisian French, between the ages of 20 and 27 years (Mean = 23 years), and whose parents were also speakers of Parisian French. All participants were university students, or recent graduates who had not lived outside France for more than six months. They were all tested in Paris by a native speaker.

4.2 Stimuli

To reduce potential variability from phonological factors other than prosody, the stimuli consisted of 20 real words of the form of CV 'CVCLə, where "L" is a liquid, and the preceding C is a voiced obstruent, e.g., *lavable* [la'vablə] 'washable', *cadavre* [ka'davrə] 'corpse'; stress is on the penultimate syllable, as noted. To increase the chances of schwa production, the stimuli appeared in a sentence before the word *cet* 'that.' So, the sequence of the three consonants (CL_ #[s]) created a LTC context where the second consonant is also the sonority peak of the sequence (i.e., satisfying Côté's (2001) condition for schwa production).

Short dialogues were used to elicit the targets in two prosodic conditions (Figure 1). In the Focus Condition, the target words appeared in the focused position of a sentence, where there was an Intonational Phrase boundary to the right (3a). In the Non-Focus Condition, a word following the target was focused, such that the target was followed by a weaker boundary, which we assume here is a Phonological Phrase boundary (3b).

(3) Prosodic boundaries in Focus and Non-Focus Conditions.

Note that the target is *buvable* 'drinkable' and the focused word is bolded.

a. Target focused: *Marie a dit «**buvabl**[ə]»]$_{IPh}$ cet après-midi.*
'Marie said "**buvable**" this afternoon.'

b. Target not focused: *Marie a dit «buvabl*[ə]»]$_{PPh}$ *cet **après-midi,** pas ce matin.*
'Marie said "buvable" this **afternoon**, not this morning.'

4.3 Procedure

Following an introduction and practice trials, the participants proceeded to the actual experiment. They were presented with a series of PowerPoint slides where they saw two people in conversation (Figure 1). They were instructed to read both parts of each dialogue in a natural way; bolding indicated which word was focused, to help elicit the intended focus condition.

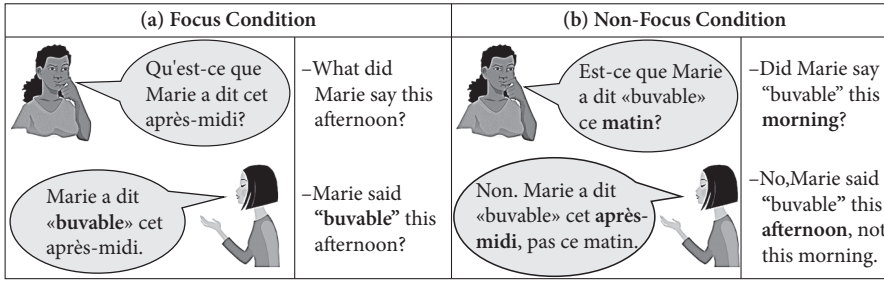


Figure 1. Examples of the two focus conditions

The dialogues were presented in two pseudo-random orders, and the speech was recorded with a head-mounted microphone in Praat.

4.4 Analyses

The data were coded in three categories, illustrated in Figures 2a–c: (a) CLə = the schwa is clearly present (e.g., [lavablə]; Figure 2a); (b) C = the schwa and the

liquid are clearly absent (e.g., [lavab]; Figure 2b); (c) CL = unclear syllabic material following the obstruent but with no clear final schwa (e.g., [lavabl]; Figure 2c). The percentage of words in each category was calculated.

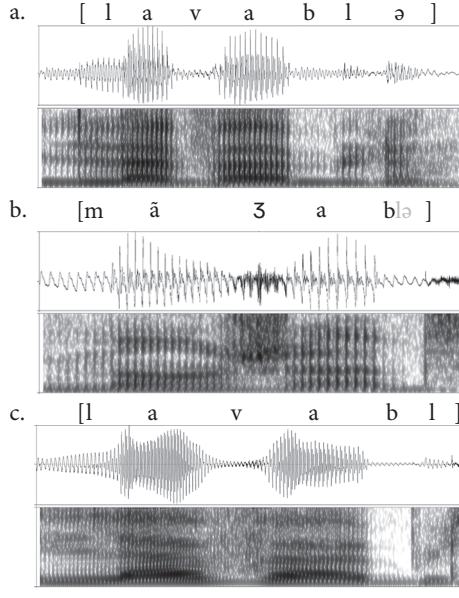


Figure 2. Examples of the three coding categories

5. Results

First, considering all of the speakers' responses together, we find that in the Non-Focus condition, 63% of the words were produced with a schwa; however, we observed two main strategies among the participants. The first strategy (A) was to produce the final schwa, or at least some syllabic material, in the expected position, and the second strategy (B) was to remove both the final schwa and the preceding liquid. As Figure 3 shows, six speakers used Strategy A, since they produce the liquid and the final schwa in 76%–100% of the words. The other three speakers used Strategy B, deleting both the liquid and the final schwa in 84%–100% of the words.

Moving to the Focus condition (Figure 4), there is an increase in schwa production, with 74% of the schwas produced overall, in comparison to the Non-Focus results. The speakers who used Strategy A did not change their behavior. That is, they still produced the liquid and final schwa most of the time, with a

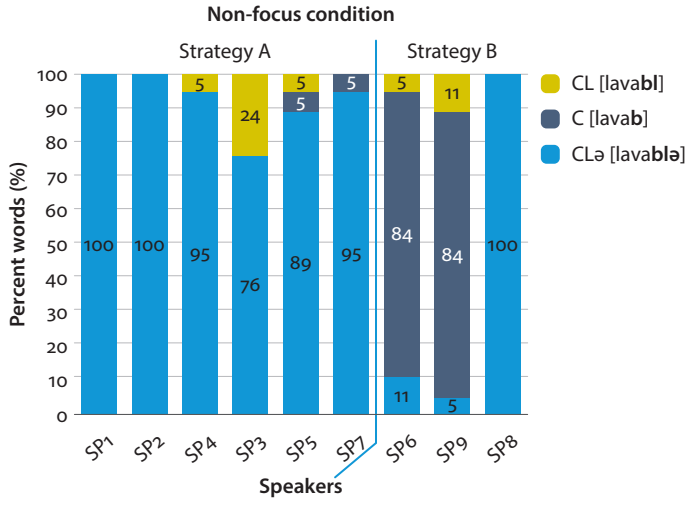


Figure 3. Results per speaker in the Non-Focused condition

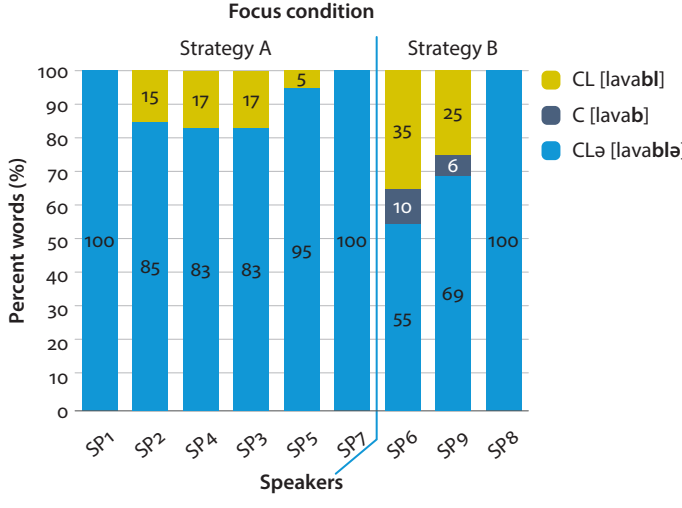


Figure 4. Results per speaker in the Focused condition

modest increase under Focus (83%–100%). On the other hand, the Strategy B speakers changed in the Focus condition. With the exception of Speaker 8, who deletes the liquid and the schwa in both focus conditions, the other Strategy B speakers rarely deleted the last syllable (6%–10%) when the word was focused, producing the final schwa in more than half of the words (55%–69%).

6. Discussion

In this paper, we have investigated the distribution in the production of word-final schwa in Parisian French, including a systematic analysis of the effect of prosodic boundaries on the production. While final schwa in Parisian French is typically not produced, it may be produced when a word ends with a consonant cluster and is followed by a word that starts with a consonant (*loi des trois consonnes*). There is, however, considerable variability in this context. As is well known, some of this variability comes from phonological, dialectal, stylistic, etc. factors. We thus controlled the non-phonological factors to the extent possible, in order to specifically test the effect of the prosodic structure on schwa production, as well as its variability.

As seen in (3), repeated here in (3′), we compared the effects of two prosodic boundaries, a stronger (IPh) boundary and a weaker (PPh) boundary. We found that schwa production was greater at an IPh prosodic boundary than at a PPh boundary, corresponding to our Focus and Non-Focus Conditions, respectively.

(3′) Prosodic boundaries to the right of the target word for each Focus Condition: Schwa Production
   a. *Marie a dit «**buvabl**[ə]»]$_{\text{IPh}}$ cet après-midi.* (Foc): schwa = 74%
   b. *Marie a dit «buvabl*[ə]»]$_{\text{PPh}}$ *cet **après-midi**.* (Non-Foc): schwa = 63%

The results from this study confirm Hypothesis 2, and thus support Hansen's (2012) position to the effect that schwa is more likely to be observed with stronger boundaries; and as such, they disconfirm Hypothesis 1, corresponding to Côté's (2001) claim that schwa is more likely to appear with weaker boundaries.

It must be noted, however, that there are important differences between the type of schwa considered in Hansen (2012) and Côté (2001). That is, Hansen (2012) examines prepausal schwas (i.e., word-finally before a pause), independently of their etymological, and potentially underlying, status. Côté (2001), on the other hand, discusses only etymological schwas, and these may occur in any position in the word (initial, medial, final); the comparison with our findings therefore really only pertains to the final schwas. If we compare, the percentages of schwa production in Hansen (2012) for the prepausal (18%–26%) and non-prepausal schwa (67%–84%), we see that the schwa production in our Non-Focus condition (63%–74%) is quite similar to Hansen's non-prepausal schwa production, that is, much more frequent than the schwa production in the prepausal position. We do not actually observe differences in the pauses (cessation of phonation) in our two focus conditions, so we can conclude more precisely that the differences in schwa production that we observe are, in fact, due to the prosodic constituent structure, not pausing. With regard to Côté's (2001) analysis, we must note that it

is largely based on the results from Dell's (1977) experiment, and her own intuitions on relevant phrases. Examination of Dell's data shows, moreover, that only a few segmental contexts were tested, and they involved voiceless obstruents or /r/+obstruent sequences, so the difference with our results may, in fact, be due to different segmental contexts. It is also possible that changes have taken place in the language between 1977 and 2012, resulting in somewhat different patterns of schwa production.

Another noteworthy finding in our study was the different production patterns between the two groups of speakers we identified. That is, in one group, the speakers produced schwa the majority of the time at the weaker PPh boundary, and thus showed minimal change at the stronger IPh boundary. That is, since the rate of schwa production was very high to start with, there was not much possibility for a change at the IPh boundary in the Focus condition. The big difference we observed was in the speakers who dropped schwa at the PPh boundary, and then showed a much higher rate of schwa production at the stronger IPh boundary. Thus, we see that while prosody does affect schwa production, its effect appears to differ systematically between two groups of speakers.

In addition to the presence or absence of schwa, we found that we needed to actually distinguish three main behaviors: (a) production of the full syllable, (b) deletion of the liquid and schwa, and (c) production of the liquid without the schwa (Figure 1 above). Three of the speakers exhibited all three behaviors across all the words, while four speakers exhibited only two and two speakers exhibited only a single behavior at the two extremes: 100% schwa production, and a similarly consistent deletion not only of the schwa, but also the preceding liquid. It is thus clear that in order to understand schwa variability among speakers, we must not only consider two choices, presence or absence of schwa, but also a third option, where some sonorant material appears in the place of the expected schwa, but it does not exhibit a clear liquid + schwa sequence.

Thus, we find that our results support Hypothesis 3, and generally confirm Bayles et al.'s (2016) findings that schwa production is variable within and across speakers, even when controlling for known factors that introduce variability. As Bayles et al. (2016) point out, this raises the interesting challenge of accounting for variability in the phonological system of the language, whether in a rule-based or constraint-based model.

7. Conclusions

In sum, we have investigated the effects of prosodic structure on the production of word-final schwa in Parisian French, and the variability within and across

speakers. To isolate the prosodic effects, we controlled for other factors that might affect the production of schwa (segmental context, dialectal, social class, speech style). To elicit final schwas, we used words ending in CC clusters, and preceding another consonant at the beginning of the next word, a context especially favorable for schwa production (*loi des trois consonnes*). We found that most of the speakers did, in fact, exhibit considerable final schwa production, even though Parisian French is a dialect that typically deletes it. This behavior was not uniform, however, with some speakers always or mostly producing the schwa, and others always or mostly deleting it. The prosodic effect of the stronger prosodic boundary was most evident in the speakers with lower schwa production in general, but overall the pattern that clearly emerged was an increase in schwa production at a stronger prosodic boundary.

References

Bayles, A., A. Kaplan, and A. Kaplan. 2016. "Inter- and Intra-speaker Variation in French Schwa." *Glossa: A Journal of General Linguistics* 1 (1): 19, 1–30. https://doi.org/10.5334/gjgl.54

Carton, F. 1999. "L'épithèse vocalique en français contemporain: étude phonétique." *Faits de Langue* 13: 35–45. https://doi.org/10.3406/flang.1999.1236

Côté, M.-H. 2001. "Schwa Deletion and Epenthesis in French." In *Consonant Cluster Phonotactics: A Perceptual Approach*, ed. by M.-H. Côté, 79–133. Cambridge, MA: MIT dissertation.

Côté, M.-H. 2007. "Le rôle des représentations sous-jacentes et des contraintes de surface dans la distribution de différentes catégories de schwa en français." In *Schwa(s): 5e Journées d'études linguistiques de Nantes*, 205–210. Université de Nantes.

Dell, F. 1985. *Les règles et les sons*. Coll. Savoir, Paris: Hermann.

Dell, F. 1977. "Parametres syntaxiques et phonologiques qui favorisent 1'epenthese de schwa en franqais moderne." In *Actes du Colloque franco-allemand de linguistique thorique*, ed. by C. Rohrer, 141–153. Tiibingen: Niemeyer. https://doi.org/10.1515/9783111681351-008

Eychenne, J. 2005. "The Emergence of Prosodic Schwa in French." *Presentation at Bilbao-Deusto Student Conference in Linguistics*.

Eychenne, J. 2006. "Aspects de la phonologie du schwa dans le français contemporain: Optimalité, visibilité prosodique, gradience." Université de Toulouse-Le Mirail dissertation.

Eychenne, J. 2009. "La pronunciation du français au Pays Basque." In *Phonologie, variation et accents du français*, ed. by J. Durand, B. Laks, and C. Lyche, 231–258. Paris: Hermès.

Eychenne, J. 2015. "Observations on the Phonetic Realization of Opaque Schwa in Southern French." *Studies in Phonetics, Phonology and Morphology* 21 (3): 457–494. https://doi.org/10.17959/sppm.2015.21.3.457

Fónagy, I. 1989. "Le français change de visage?" *Revue Romane* 24(2): 225–254.

Gess, R., C. Lyche, and T. Meisenburg 2012. *Phonological Variation in French: Illustrations from Three Continents*. Amsterdam: John Benjamins. https://doi.org/10.1075/silv.11

Hansen, A. B. 1994. "Etude du E caduc – stabilisation en cours et variations lexicales." *Journal of French Language Studies* 4: 25–54. https://doi.org/10.1017/S0959269500001964

Hansen, A. B. 1997. "Le nouveau [ə] prépausal dans le français parlé à Paris." In *Polyphonie pour Iván Fónagy*, ed. by J. Perrot, 173–198. Paris: L'Harmattan.

Hansen, A. B. 2012. "A Study of Young Parisian Speech: Some Trends in Pronunciation." In *Phonological Variation in French: Illustrations from Three Continents*, ed. by R. Gess, C. Lyche, and T. Meisenburg, 152–172. John Benjamins. https://doi.org/10.1075/silv.11.09han

Hansen, A. B., and M.-B. M. Hansen 2003. "Le [ə] prépausal et l'interaction." In *Structures linguistiques et interactionnelles dans le français parlé*, ed. by A. B. Hansen, and M.-B. M. Hansen, 89–109. Copenhagen: Museum Tusculanum Press.

Lacheret, A., and C. Lyche 2008. "Looking at French Schwa through the Glasses of Prosody." *Proceedings of Speech Prosody 2008*.

Léon, P. R. 1993. *Précis de phonostylistique. Parole et expressivité*. Paris: Nathan Université.

Lucci, V. 1983. *Etude phonétique du français contemporain à travers la variation situationnelle: Débit, rythme, accent, intonation, ə muet, liaisons, phonèmes*. Université des langues et lettres de Grenoble.

Pulgram, E. 1961. French /a/: Statics and Dynamics of Linguistic Subcodes." *Lingua* 10: 305–325. https://doi.org/10.1016/0024-3841(61)90109-7

Pustka, E. 2009. "PFC et la phonologie du français en Vendee." In *Phonologie, Variation et accents du français*, ed. by J. Durand, B. Laks, and C. Lyche, 307–335. Paris: Hermès.

Tranel, B. 2000. "Aspects de la phonologie du français et la théorie de l'optimalité." *Langue Française* 126: 39–72. https://doi.org/10.3406/lfr.2000.989

Vogel, I., A. Athanasopoulou, and N. Pincus 2016. "Prominence, Contrast and the Functional Load Hypothesis: An Acoustic Investigation." In *Dimensions of phonological stress*, ed. by J. Heinz, R. Goedemans, and H. van der Hulst, 123–167. Cambridge: University Press. https://doi.org/10.1017/9781316212745.006

Walter, H. 1990. "Une voyelle qui ne veut pas mourir." In *Variation and Change in French. Essays Presented to Rebecca Posner on the Occasion of her Sixtieth Birthday*, ed. by J. Green, and W. Ayres-Bennett, 27–36. London and New York: Routledge.

Agreement and pronouns

Implications for partial control

J.-Marc Authier & Lisa A. Reed
The Pennsylvania State University

This chapter compares the workings of agreement mismatches involving the French pronoun *on* with those involving PRO in partial control (PC) environments in order to advance our understanding of the type of plurality that has been assumed to be associated with the latter. The theoretical framework used to effect this comparison is the two-step theory of Agree argued for by Arregi and Nevins (2012) and Smith (2017). The evidence uncovered suggests that whatever is assumed to induce the plurality of PC PRO (e.g. Landau's (2016b) associative morpheme) leaves neither morphological nor semantic agreement footprints. We also show that, on the LF side, the type of plural set denoted by PC PRO appears to be immune to any sort of quantificational manipulation. We are thus left with an inference of plurality with no tangible grammatical correlates, which points in the direction of a process of pragmatic associative reference along the lines of Haug (2014).

Keywords: agree, partial control, plurality, secondary predicates, semantic agreement

1. Introduction

Authier and Reed (2018) present evidence that only covert lexical reciprocals, which are symmetric in the sense spelled out in (1) can be predicated of Partial Control (PC) PRO.

(1) A reciprocal predicate is symmetric if it expresses a relation between participants that is not based on the accumulation of sub-events but is instead based on an atomic event.

According to (1), a predicate like *s'embrasser* 'kiss' is non-symmetric in that it gives rise to event-counting ambiguities of the sort illustrated in (2); that is, it can be interpreted as the accumulation of sub-events, as shown in (2ii).

https://doi.org/10.1075/rllt.16.02aut

(2) *Alain et Amélie se sont embrassés trois fois.*
Alain and Amélie SE are kissed three times
'Alain and Amélie kissed each other three times.'
i. There was a total of three kissing events.
ii. There were six kissing events; i.e., Alain kissed Amélie three times and she kissed him back three times.

A predicate like *se réunir* 'meet', on the other hand, is said to be symmetric in that it can only denote atomic events, as (3) illustrates.

(3) *Alain et Amélie se sont réunis trois fois.*
Alain and Amélie SE are met three times
'Alain and Amélie met three times.'
i. There was a total of three meeting events.

What Authier and Reed (2018) show is that only those reciprocal verbs that have symmetric semantics are compatible with PC PRO. This constraint on PC is thus responsible for contrasts such as the one in (4).[1]

(4) a. *Alain$_j$ préférait [PRO$_{j+}$ se réunir dans le salon].*
Alain preferred SE meet in the lounge
'Alain preferred to meet in the lounge.'
b. **Alain$_j$ préférait [PRO$_{j+}$ s'embrasser dans le salon].*
Alain preferred SE-kiss in the lounge
'Alain preferred to kiss in the lounge.'

Authier and Reed (2018) go on to argue that the theory of PC proposed by Landau (2016a, b), which takes PRO in PC to be a group-denoting, syntactically singular but semantically plural noun does not predict the fact that it can only be the subject of symmetric predicates. Indeed, there are two sets of facts discussed in the literature that show that PC PRO and group-denoting nouns exhibit divergent behaviors.

The first distinction between group nouns and PC PRO was pointed out in Pearson (2013: 312): In many English dialects spoken in the UK, group-denoting nouns (a.k.a. collective nouns) can bind plural anaphors in the so-called periphrastic reciprocal construction, as in (5), and are, therefore, compatible with non-symmetric reciprocals. This, however, does not correlate with PC PRO being able to antecede plural anaphors in those dialects (cf. (6)), as one might expect.

1. Note that English *kiss*, unlike its periphrastic counterpart *kiss each other*, is, according to the definition in (1), symmetric, hence the grammaticality of the English gloss in (4b) is expected.

Thus, within the same English dialect, we see that group nouns, but not PC PRO, can function as the subject of a non-symmetric reciprocal predicate.[2]

(5) a. %With this new policy, the government have offered each other up for criticism.
b. %The team hate each other.

(6) *The platoon leader$_j$ would prefer [PRO$_{j+}$ to consult each other before the assault].

A second distinction between group nouns and PC PRO is introduced in Authier and Reed (2018), who show that the two differ with respect to their (in)compatibility with those non-symmetric reciprocal predicates that are not periphrastic reciprocals. PC PRO can only be predicated of symmetric reciprocal predicates; that is, predicates that *always* yield an atomic event interpretation. Group nouns, on the other hand, are not subject to this constraint: they are compatible with non-symmetric reciprocal predicates on both of their interpretations, as shown in (7). The predicate *s'envoyer des cartes postales* 'send postcards to each other' in (7b) is a non-symmetric reciprocal predicate that takes the group noun *couple* as its subject and is incompatible with an atomic event interpretation (i.e. it can only be understood as the accumulation of postcard sending sub-events). In (7a), on the other hand, the non-symmetric reciprocal *s'embrasser* 'kiss', predicated of the same group noun, is compatible with both an accumulation of sub-events and an atomic event interpretation. Thus, group nouns are unlike PC PRO in that they are compatible with non-symmetric predicates.

(7) a. *Le couple s'est embrassé toute la journée.*
the couple SE-is kissed all the day
'The couple kissed (each other) all day long.'
b. *Le couple s'est envoyé des cartes postales tout l'été.*
the couple SE-is sent some postcards all the-summer
'The couple sent each other postcards all summer long.'

Our goal in the present chapter will be to carefully examine a second type of issue tied to Landau's characterization of PC PRO, one that has to do with the mismatch between the morphological and semantic person values on PC PRO made evident by sentences like (8).

---

2. Landau (2000: 50) give examples similar to (6) as grammatical (in the relevant dialects), which indicates that there may be some dialectal variation. The existence of the judgments reported by Pearson (2013: 312), confirmed independently by Authier & Reed (2018: 384), is, however, enough to discredit the theory that PC PRO always behaves like a group-denoting noun.

(8) a. *Le sujet sur lequel je$_j$ voudrais [PRO$_{j+}$ m$_j$'entretenir]*
the topic on which I would-like SE-1SG-converse
est grave.
is serious
'The matter I would like to discuss is a serious one.'

b. *Préfèrerais-tu$_j$ [PRO$_{j+}$ te$_j$ rassembler] autre part?*
would-like-you SE-2SG assemble elsewhere
'Would you prefer to assemble elsewhere?'

In (8a), the [1st person SG] form of the reciprocal morpheme *me* in the infinitive indicates that PC PRO is morphologically [1st person SG] and thus matches the morphological number value of the controller *je* 'I' in the matrix. Semantically, however, PC PRO in (8a) is understood as denoting a [1st person PL]. Similarly, in (8b), PC PRO is morphologically [2nd person SG], as evidenced by the *te* form of the reciprocal morpheme, but it is semantically [2nd person PL]. Landau (2016b) explains such mismatches as follows. First, he assumes that PC constructions contain an abstract associative morpheme (AM) attached to the inflectional head of the infinitive containing PRO. This abstract morpheme functions as a group operator on the index of the controller, which yields a set that includes the referent of this index plus at least one other referent. Because the abstract morpheme is attached to a phi-less T, it does not induce morphological plurality, though it does induce semantic plurality. Finally, while PC PRO inherits its morphological ϕ-features from the controller, it does so at PF, which is "too late" to impact the semantic interpretation. While Landau's PF account does make the correct predictions in this particular case, it, in turn, raises some non-trivial questions regarding the nature of agreement and that of PRO. In this chapter, we will examine these questions in some detail and will assess their impact on possible theories of PC.

2. Agreement mismatches with pronouns

PRO is standardly assumed to be a phonologically silent pronoun. It therefore stands to reason that we should seek to find out if there are phonologically realized pronouns that exhibit agreement mismatches similar to those exhibited by PRO in PC infinitivals. An obvious candidate is so-called "first person plural *on*", a pronoun that substitutes in colloquial French for the pronoun *nous* 'we', the latter being confined to prescriptive/high register French. That is, this type of *on* is morpho-syntactically [3rd person SG] but, semantically, it refers to a group that

includes the speaker.[3] This dichotomy is reflected by the fact that 1st person plural *on* triggers [3rd person SG] agreement on the verbal element in T and the choice of the [3rd person SG] form of the reflexive or reciprocal morpheme *se*, but in all other respects it triggers the choice of [1st person PL] (or just plural) forms; for example, on the adjective embedded under predicative *être* 'be' in (9a), on the emphatic anaphor it binds in (9b), and on the secondary adjectival predicate it is construed with in (9c).

(9) a. *On a été* ***loyaux****.*
ON has been loyal-PL
'We've been loyal.'

b. *Si tu veux pas nous filmer, on se filmera* ***nous-mêmes****.*
if you want not us to-film ON SE will-film ourselves
'If you don't want to film us, we'll just film ourselves.'

c. *On est parti de là* ***joviaux*** *comme des*
ON is left from there jovial-PL like some
titis parisiens.
rascals Parisian
'We left that place jovial like Parisian street urchins.'

Beginning with Morgan (1984), "hybrid" nominals, that is, nominals that allow variable agreement options that target either their morphological shape (e.g., 1st person SG for *on*), or their semantic interpretation (e.g., 1st person PL for *on*), have led to the recognition of both morphologically-based and semantically-based agreements. This has, in turn, led to new formulations of AGREE that assume that AGREE is a two-step process (Benmamoun et al. 2009; Arregi & Nevins 2012;

3. This is by no means the only definite interpretation of *on* since this pronoun also has what Creissels (2011) calls a de-personalizing use, often characterized as "stylistic" in French grammar books. In this use, *on* can be replaced by a second person pronoun (i.e., *tu* 'you-SG' or *vous* 'you-PL') without changing the truth value of the sentence in which it is found. This type of *on* conveys a condescending attitude towards the referent, hence it typically presupposes a hierarchical relation whereby the speaker considers himself/herself superior to the addressee in some way, as the following examples illustrate.

(i) *Eh toi! Le rouquin au deuxième rang! On fait le malin?*
hey you the redhead in-the second row ON makes the smart-ass
'Hey you! The redhead in the second row! Are you being a smart-ass?'

(ii) *Alors, ma petite, on est remise de ses émotions?*
so my little-one ON is recovered from her emotions
'So, young lady, did you manage to pull yourself together yet?'

Smith 2017) and that ϕ-features are composed of two halves: a *u*F corresponding to a formal morphological value and an *i*F corresponding to a semantic value (see Smith 2015, 2017 and Wurmbrand 2017). Under this view, a hybrid pronoun like *on* embodies a divergence of values on the *u*F and *i*F of its number feature (i.e., [*u*F:SG, *i*F:PL]). The two number feature halves are paired together in the narrow syntax before being separated at the point of transfer to the PF and LF components. For concreteness, let us take the version of AGREE proposed in Smith (2017), who, building on the work of Arregi & Nevins (2012), defines AGREE as in (10).[4]

(10) Agreement by Probe with Goal proceeds in two steps:
   a. AGREE-LINK: A probe has unvalued ϕ-features that trigger Agree with a goal (possibly more than one). The result is a link between probe and goal.
   b. AGREE-COPY: After the syntactic derivation, one of the values of the ϕ-features of the goal is copied onto the probe linked to it by AGREE-LINK. If AGREE-COPY occurs at the point of transfer, the goal must c-command the probe.

Smith (2017) further assumes that if AGREE-COPY chooses to take its value from the *i*Fs of the goal, it can only do so at the point of transfer, that is, right before features are sent to the interfaces. This allows *i*Fs to play a role in (phonologically overt) agreement. If, on the other hand, AGREE-COPY takes its value from the *u*Fs of the goal, it can do so either at the point of transfer or in the post-syntactic PF component. Thus, when AGREE-COPY targets *u*Fs, there is no requirement that the goal c-command the probe; that is, what we called morphological agreement is an operation that, unlike semantic agreement, can look "upwards" or "downwards" in the structure. Finally, whether AGREE-COPY operates on *u*Fs or *i*Fs with respect to a particular agreement target is arbitrarily decided by each language or dialect of a language, although, as first noted by Corbett (1983), there is a cross-linguistic tendency for expressions in closer "proximity" to nouns to show morphological agreement whereas those more distant tend to show semantic agreement.[5] For example, in (9b), the subject pronoun *on* is a goal endowed with [*u*F:SG, *i*F:PL]. Its closer probes, namely, the reflexive morpheme and the tensed verb in T use its [*u*F:SG] for AGREE-COPY, yielding [1st person SG] forms for both elements, but the more distant probe that is the emphatic reflexive copies the [*i*F:PL] feature of the same probe, yielding the [1st person PL] form *nous-mêmes*

4. Note that outside the generative literature on agreement, the probe is standardly called the "target" and the goal is referred to as the "controller".

5. See also Kathol (2000, Section 4.1) for an attempt to derive Corbett's agreement hierarchy.

'ourselves'. That this is just a cross-linguistic tendency is, however, shown by the difference between American English and British English regarding the features selected by COPY-AGREE applying between a T probe and a collective noun subject goal. As seen in (11), British English, but not American English, is more likely to violate the "proximity" generalization just discussed by allowing COPY-AGREE to use the [*i*F:PL] of the goal rather than its [*u*F:SG], yielding (11b).

(11) a. Our company's legal team is investigating the incident.
b. Our company's legal team are investigating the incident.

Keeping the tenets of Smith's (2017) theory of Agree in mind, let us now turn to the kind of agreement exhibited by depictive adjectival secondary predicates in French.

3. Agree and syntactically-dependent depictive secondary predicates

Depictive adjectival secondary predicates with external argument modification such as the ones in bold in Example (12) denote a state that holds true of the nominal they modify during the entire duration of the event denoted by the sentence in which they appear, which accounts for the fact that they do not usually involve individual-level predicates since the latter denote states that are independent of any eventuality (13).[6] In other words, depictive secondary predicates are concealed time interval predicates. Thus, (12a) states that Monique was in a state of drunkenness that held true during the entire span of the past singing event she was involved in. Similarly, (12b) does not ascribe to the soldiers the property of being exhausted/soaked as a final state resulting from the mountain-crossing event in which they were involved but states instead that this property held throughout the mountain-crossing event in which they were participants.

---

6. However, as (i) shows, individual-level predicates like *bilingual* (viz. **There were two students bilingual*) denote properties that can be developed (or lost) during an individual's lifetime and are therefore possible as depictive secondary predicates when the sentence references a transition/change of state (cf. Ardid-Gumiel 2001).

(i) *Cédric est allé à l'Ecole hôtelière de Lausanne et en est sorti bilingue.*
Cédric is gone to the-school hospitality of Lausanne and from-it is come-out bilingual
'Cédric attended the Lausanne School of Hotel Management and came out of it bilingual.'

(12) a. *Monique a chanté* ***soûle***.
Monique has sung drunk-FEM.SG
'Monique sang drunk.'

b. *Les soldats ont franchi les*
the soldiers have gone-over the
montagnes ***épuisés/trempés***.
mountains exhausted/soaked-MASC.PL
'The soldiers crossed the mountains exhausted/soaked.'

(13) a. **Monique a chanté jolie*.
Monique has sung good-looking-FEM.SG
'*Monique sang good-looking.'

b. **Les soldats ont franchi les montagnes malins/grands.*
the soldiers have gone-over the mountains clever/tall-MASC.PL
'*The soldiers crossed the mountains clever/tall.'

The question of what kind of syntactic structure is best-suited to derive the semantic and syntactic properties of subject-oriented depictive secondary predicates is still being debated in the literature. The oldest proposal is Stowell's (1981) right-adjoined small clause with PRO analysis, illustrated in (14), which reflects more recent assumptions regarding phrase structure.

(14) $[_{vP}$ *Monique* $[[_{VP}$ *chanté*$]$ $[_{SC}$ PRO *soûle*$]]]$

This analysis, often referred to in the literature as the "standard analysis", has been assumed by a number of authors since (as recently as Pylkkänen 2008). The alleged presence of PRO in depictive secondary predicate constructions has, however, been argued by Marušič, Marvin and Žaucer (2003) to be problematic, based on case-marking data from Slovenian. Additionally, Ardid-Gumiel (2001) and Bruening (2018) have argued on semantic grounds that depictive constructions should be given a monoclausal analysis and that therefore, it should not be assumed that the depictive adjective and the nominal it modifies form a small clause, since this would presumably imply bi-clausality. On the other hand, Ardid-Gumiel (2001: 17) presents evidence from Spanish indicating that the argument the depictive adjective is predicated of is first merged as the specifier of the depictive adjective (see also Marušič, Marvin & Žaucer 2003 and Irimia 2005 for similar ideas). In particular, she notes that floating quantifiers like *todos* 'all' in Spanish can appear right before the depictive adjective. We illustrate the phenomenon with French *tous* 'all' in (15).

(15) a. *Les enfants sont arrivés tous épuisés mentallement.*
the children are arrived all-PL exhausted-PL mentally
'The children arrived all mentally exhausted.'

b. *Les enfants ont planté leurs fleurs tous trempés*
the children have planted their flowers all-PL soaked-PL
jusqu'aux os.
to-the bones
'The children planted their flowers all soaking wet.'

Her analysis, which we adopt here, makes use of sideward movement (cf. Nunes 2001, 2004 and Hornstein 2001 among others). Sideward movement, like standard movement, is an instance of Merge. However, sideward movement is a sequence of derivational steps that consists in copying a constituent α contained in a syntactic phrase K and merging this copy of α with another syntactic phrase L, which has been independently constructed and is, at this point, unconnected to K. Merging α with L results in a new phrase marker M. M and K are then merged to form a syntactic object. Accordingly, the (partial) derivation for a sentence like (12a) proceeds as in (16).

(16) a. [$_{v\text{P}}$ *v chanté*], [$_{\text{AP}}$ *Monique soûle*]
→ The vP and the AP phrase markers are constructed independently.

b. [$_{v\text{P}}$ *Monique* [*v chanté*]], [$_{\text{AP}}$ <*Monique*> *soûle*]
→ (External) merger of *Monique* with the *v*P projection, forming a *v*P specifier (sideward movement of *Monique*)

c. [[$_{v\text{P}}$ *Monique* [*v chanté*]] [$_{\text{AP}}$ <*Monique*> *soûle*]]
→ Merger of the AP with the *v*P
(The derivation then proceeds to construct the rest of the clause.)

As shown in (16a), the depictive adjective *soûle* 'drunk' and the nominal phrase it modifies (*Monique*) merge to form an AP while the main verb *chanté* 'sung' and the *v* merge independently to form a *v*P. The next step, shown in (16b), involves an instance of sideward movement, which consists in copying *Monique* in the AP and merging it with the *v*P, forming a specifier position of *v*. Finally, as shown in (16c), the two independently constructed subtrees merge, forming a vP constituent where *Monique* is both the argument of the depictive adjective and the external argument of the verb. Once T is merged in the structure, its EPP feature will then trigger internal merge of *Monique*, forming a specifier of T in the usual way. If sentential negation (*pas* 'not') is merged prior to this, as in (17), it will c-command the AP and thus, scope over it, correctly predicting that (17) lacks the reading where Monique, in a state of inebriation, didn't sing (cf. Roberts 1988).

(17) [$_{\text{TP}}$ *Monique a* [$_{\text{NegP}}$ *pas* [[$_{v\text{P}}$ <*Monique*> [*v chanté*]]
Monique has not sung
[$_{\text{AP}}$ <*Monique*> *soûle*]]]]]
drunk
'Monique did not sing drunk.'

4. On some agreement puzzles for PC PRO

Having established a plausible syntactic structure for depictive secondary predicate constructions, we turn, in this section, to the kind of agreement depictive adjectives display in such constructions. We will then pay special attention to those that involve agreement mismatches and PC PRO.

4.1 Floated quantifiers

As the reader will recall, so-called first person plural *on* is a goal endowed with [*u*F:SG, *i*F:PL]. In (18), its closest probes, namely, the morpheme *se* and the tensed verb *rappelle* 'remember', use its [*u*F:SG] for AGREE-COPY, yielding [3rd person SG] forms for both elements. The example in (18), in which *on* is the controller of PRO, additionally shows that as a probe under obligatory control, PRO uses the [*u*F:SG] of its controller for AGREE-COPY, as evidenced by the [3rd person SG] form of the reciprocal morpheme *se* on the embedded auxiliary. Prior to AGREE-COPY, however, PRO in (18) has undergone AGREE-LINK with its controller, which means that it, like its controller, is endowed with [*i*F:PL]. This plurality, however, does not impact AGREE-COPY and thus remains confined to interpretation. Importantly, this is sufficient for the floated quantifier *tous* 'all' to be compatible with PRO in this case.

(18) *On*$_j$ *se rappelle [PRO*$_j$ **nous/s'être tous embrassés à*
we remember *1PL/3SG-be all kissed at
cette occasion].
that occasion
'We remember us all kissing on that occasion.'

What is interesting about the status of PRO in (18) is that it closely resembles PC PRO in that it bears the singular morphological feature of the controller in the matrix, yet it is semantically interpreted as a group that includes the speaker, hence its compatibility with *tous* 'all', which quantifies over the members of that group. Unexpectedly, however, PC PRO, though it is standardly assumed to semantically denote a group that includes the speaker in a sentence like (19), turns out to be incompatible with *tous*.

(19) *Je*$_j$ *me rappelle [PRO*$_{j+}$ *m'être (*tous) réuni dans la*
I remember 1SG-be (*all) met in the
salle de séminaire]
room of seminar
'I remember (all of us) meeting in the seminar room.'

Assuming Landau's theory of PC, the only difference between (19) and (18) is that in (18), PRO inherits its semantic plurality directly from the controller, whereas in (19), the semantic plurality of PRO is induced by the silent associative morpheme on the infinitival T. Why this should make a difference with respect to the availability of universal quantification via *tous* is not immediately obvious, however. What is clear is that the semantic notion of plurality ascribed to PC PRO by Landau and others is insufficient to account for the contrast between (18) and (19). Note incidentally that the fact that the reference of PC PRO does not appear to be exhaustively determined by its syntactic controller cannot be taken as reliable proof of its *semantic* plurality. By this we mean that upon hearing a sentence like (20a), in which the reference of PRO *is* exhaustively determined by the controller, one draws the inference that the speaker and other undetermined individuals met. This inference appears to involve something akin to notional plurality, which is based on our reasoning about parts and wholes based on our knowledge of the world; that is, one cannot attend a meeting event on one's own since meetings are, by definition, gatherings of people. Interestingly, as (20b) shows, this type inference turns out to also be unable to introduce in the semantics a plural set over which *tous* can quantify.

(20) a. *Je$_j$ me rappelle [PRO$_j$ avoir assisté à la réunion].*
I remember to-have attended at the meeting
'I remember attending the meeting.'

b. **Je$_j$ me rappelle [PRO$_j$ avoir tous assisté à la réunion].*
I remember to-have all attended at the meeting
'I remember us all attending the meeting.'

One might, of course, object that collective predicates like *se réunir* 'meet' in (19) apply only to collections/non-atomic individuals and that this can be taken as proof that PRO in (19) is semantically plural. Let us not forget, however, that the availability of PC PRO is not determined by its being the subject of a collective predicate alone. It is also contingent upon its being embedded under a matrix non-simultaneous attitude predicate (Pearson 2013) and its being the subject of a symmetric predicate (Authier & Reed 2018). We should also keep in mind that it is presently far from clear how those factors conspire to induce the "plurality" of PC PRO and that a full answer to this question seems necessary to determine whether this plurality is to be ascribed to the semantic or the pragmatic component (for pragmatic approaches see Bowers 2008; Duffey 2014 and Haug 2014). In what follows, we will provide additional evidence based on agreement that suggests that the plurality of PC PRO may not be semantic after all.

4.2 Agreement mismatches and the features of PC PRO

Consider first the contrast in (21). Sentences like (21a), although they do not involve PC, as (21b) does, nevertheless, mimic the latter in terms of the interpretive properties associated with the embedded subject; that is, both *on* and PC PRO end up referring to a group of individuals that includes the speaker. However, even though PC PRO and *on* are both morphologically singular and semantically plural, they differ in that only PC PRO corresponds, in (21b), to the [1st person SG] form of the reciprocal morpheme on the embedded predicate.

(21) a. *Je*$_j$ *préfèrerais* *[qu'on*$_{j+}$ **me/se* *réunisse* *demain].*
I would-prefer that-we *1SG/3SG meet tomorrow

b. *Je*$_j$ *préfèrerais* *[PRO*$_{j+}$ *me/*se* *réunir* *demain].*
I would-prefer 1SG/*3SG to-meet tomorrow
'I'd rather (we) meet tomorrow.'

This suggest that PC PRO is unlike *on* due to the fact that it strictly inherits its morphological ϕ-features from its controller, a conclusion reached by Landau (2016b). Landau further argues that only Agree-valued uninterpretable ϕ-features (i.e., *u*F/morphological features) can be stripped by Spell-Out and thus are present in PF but invisible to interpretation. This view, however, leaves us with a puzzle when we consider the type of agreement exhibited by depictive secondary predicates.

The sentence in (22) establishes that while a depictive secondary predicate linked to a goal like *on* uses its [*i*F:PL] for AGREE-COPY, tensed T probing the same goal selects instead [*u*F:SG] as the feature-half relevant to that operation.

(22) *On* *s'est* *réunis* *joviaux/*jovial.*
we 3SG-are met-PL jovial-PL/*SG
'We met in a jovial mood.'

This asymmetry in the selection of feature-halves by AGREE-COPY between depictive adjectives and T persists in the context of OC PRO controlled by *on*, as (23) shows.

(23) *On*$_j$ *se* *rappelle* *[PRO*$_j$ *s'être* *réunis* *joviaux/*jovial].*
we 3SG-remember 3SG-be met-PL jovial-PL/*jovial-SG
'We remember meeting in a jovial mood.'

Assuming with Landau (2016b) and others that PRO bears the ϕ-features transmitted by its controller and assuming further that these are composed of two halves, we must conclude that PRO in (23) inherits [*u*F:SG, *i*F:PL], undergoes AGREE-LINK with two probes – T and the depictive adjective – in the infinitival clause, and that when AGREE-COPY applies, it selects [*u*F:SG] for T (hence the [3rd person SG] form of the reflexive morpheme) and [*i*F:PL] for the depictive adjective.

With this in mind consider now the example in (24), which contains a PC PRO as well as a depictive adjective.

(24) *Eric*$_j$ *se rappelle* *[PRO*$_{j+}$ *s'être réuni jovial/*joviaux].*
Eric remembers 3SG-be met-SG jovial-SG/*jovial-PL
'Eric remembers meeting in a jovial mood.'

The sentence in (24) can only mean that Eric remembers a past meeting event in which he and other people were participants and during which *only he* was in a jovial mood. All things being equal, we expect the depictive adjective in the PC construction in (24) to select the *i*F of PC PRO. If this is the case, however, then we have to assume that PC PRO bears the feature [*u*F:SG, *i*F:SG], inherited from its controller *Eric*. But then how can PC PRO be semantically plural? To resolve this problem, Landau (2016b) assumes that in PC, AM is a syntactic affix associated with the verbal head in the infinitive that "projects a group out of the index of PRO". According to him, the shift AM induces in the semantic number of PRO is *not* the result of agreement. Rather, AM interferes with the process of indexical individuation by expanding the PRO subject of a collective predicate to a group accessed as a collective individual. What is important for our purposes is that AM is a syntactic affix that feeds the semantics of PC PRO directly rather than via its *i*F. Since, however, PC PRO inherits [*i*F:SG] form its controller in (24), this makes PC PRO simultaneously singular and plural semantically. Note that this not necessarily a contradiction given the semantics of group-nouns. Indeed, Ritchie (2014) argues that group-nouns have both a group-as-one and a group-as-many meaning. These meanings become apparent when we consider examples like those in (25).

(25) a. This committee is old. It's been around since 1789!
b. The committee had lunch in the conference room.

The sentence in (25a) illustrates the group-as-one meaning of *committee*: it does not require the members of the committee to have been around since 1789. The example in (25b), on the other hand, illustrates the group-as-many interpretation of *committee*: it requires that each member (or most members) of the committee had lunch in the conference room.

The question now becomes whether we can use this dual semantic characterization of group-nouns to arrive at a characterization of PC PRO that will derive the facts in (24). In order for the semantics of (24) to come out right, we need PC PRO to simultaneously refer to the atomic individual denoted by the controller and to some sort of group-as-one denotatum that will prevent individual members of the group from being linguistically accessible and ensure that PC PRO cannot be quantified over by *tous* (cf. (19)). The fact that PC PRO in (24) commands

singular semantic agreement on the depictive adjectival secondary predicate can be handled via the dual feature [*u*F:SG, *i*F:SG] inherited from its controller (assuming once again that COPY-AGREE selects *i*F for depictives). Additionally, the contrast between (24) and (26) underscores the necessity, under Landau's characterization of PC, to assume that the semantic plurality of PC PRO is "inactive" for agreement purposes. This is because in (26), the [*i*F:SG] (group-as-one denotatum) associated with *le groupe dont Eric fait partie* 'the group Eric is a member of' is used for COPY-AGREE by the depictive yet (26) states that *all* the members of the group (as opposed to just Eric in (24)) were jovial.

(26) *Le groupe dont Eric fait partie s'est*
the group of-which Eric makes part SE-is
*réuni jovial/*joviaux*
met jovial-SG/*jovial-PL
'The group Eric belongs to met in a jovial mood.'

We therefore come to the conclusion that even if we assume that the semantic plurality effected by Landau's AM affix yields a group-as-one denotatum, we must additionally assume, as Landau does, that it is inactive at transfer, i.e. inaccessible to either AGREE-LINK or AGREE-COPY and therefore not, in fact, akin to a group-noun. Thus, because Landau's AM leaves no syntactic or morphological footprint, one is entitled to again wonder whether the "plural" interpretation of PC PRO might fall outside the grammar proper and might instead be an instance of notional plurality i.e., a pragmatic inference involving our reasoning about parts and wholes. Indeed, it is incumbent on those purporting that PC readings arise via some grammatical device such as AM to provide empirical evidence that distinguishes a sentence like (24) from an OC example like (27), which, like (24), gives rise to the inference that Eric and other people were involved in a meeting event, does not permit universal quantification over the members of that group, and does not allow a depictive adjective to characterize this group as a whole.

(27) *Eric$_j$ se rappelle [PRO$_j$ avoir (*tous) assisté à la*
Eric remembers to-have all attended at the
*réunion jovial/*joviaux].*
meeting jovial-SG/*jovial-PL
'Eric remembers having attended the meeting in a jovial mood.'

5. Concluding remarks

In this chapter, we hope to have shown that while the interpretive contribution of PC impacts the reference of PRO so that it is not uniquely determined by the controller, neither quantification by *tous*, nor the semantic agreement found on

depictive secondary predicates is able to detect the resulting "semantic plurality" of PC PRO. In other words, what we are left with is an inference of plurality associated with PC PRO that has no quantificational or morphological correlates. Thus, pragmatic approaches to PC appear to still be viable contenders in our quest to arrive at an explanation of the phenomenon.

In the recent literature, PC is standardly defined as situations in which the reference of PRO must include that of the controller but is not exhaustively determined by it. Ironically, Lawler (1972), who is often cited as being the first to have discussed the phenomenon, took PC to be a larger phenomenon, which he described as a set of environments in which PRO stands for some organizational entity, for example, the members of the jury in (28a) (Lawler's example (14)) or the government/Air Force in (28b) (Lawler's example (15)).[7]

(28) a. That prosecutor was such a spellbinder that, if I'd been on the jury, I'd have voted to convict myself.
b. Mary wants to bomb Hanoi.

The example in (28a) would, under current assumptions, be assumed to contain a PC PRO, one that refers to a group (i.e., the jury) that includes the controller (the speaker). That is, given our knowledge of the judicial system, which crucially includes the rule that convictions can only come about as the result of a unanimous decision by the members of the jury, we understand (28a) to mean that the speaker would have voted in a way consistent with the ultimate goal of being convicted by the jury. The PRO in (28b), on the other hand, can be interpreted in one of three ways. First, it can be understood as a straightforward case of obligatory control, assuming, for example, that Mary is an Air Force pilot who wishes to go rogue and bomb Hanoi on her own. Second, (28b) appears to allow a PC reading whereby Mary, an Air Force pilot, wants for herself and other members of her squadron to bomb Hanoi. The third reading is, however, not one that would straightforwardly fall under the more recent definition of PC. On this reading, what Mary wants is a bombing by proxy; that is, she wants for the government (or, perhaps more accurately, the Air Force of her country) to bomb Hanoi on her behalf. Thus, the reference of PRO on this reading includes that of the controller only in the sense that Mary's government is her representative when it comes to bombing other nations. This is reminiscent in many ways of pragmatic associative anaphoric reference (a.k.a. bridging) as discussed in e.g. Clark & Marshall (1981: 22–23) in the context of definite descriptions like *the driver* in (29).

(29) I took a bus this morning. The driver was drunk.

7. Hence Lawler's terminological coinage of the phenomenon as "government deletion" (GD).

In (29), based on the generic knowledge that buses have drivers, the bus is a "trigger" to the "associate" driver in the sense that with the mention of a bus, people access a set of associates that constitutes a shared set of objects, hence the understood reference of the driver is the driver of the bus. Similarly, in (28b), the predicate *bomb Hanoi* acts as a trigger based on the generic knowledge that governments rather than individuals have the power to declare war and that governments represent their constituents. As a result, PRO in (28b) may acquire a proxy reading vis-à-vis the referent imposed by the controller through OC, this being a pragmatic accommodation process. We leave to future research the task to determine if similar explanations can be adduced for PC phenomena more generally.

References

Ardid-Gumiel, Ana. 2001. "The Syntax of Depictives. Subjects, Modes of Judgement and I-L/S-L Properties." In *Proceedings of the Workshop on the Syntax of Predication*, ed. by N. Zhang, 1–26. Berlin: Zentrum Allgemeine Sprachwissenschaft (ZAS).

Arregi, Karlos, and Andrew Nevins. 2012. *Morphotactics: Basque Auxiliaries and the Structure of Spellout*. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-3889-8

Authier, J.-Marc, and Lisa Reed. 2018. "Symmetric Reciprocal Semantics as a Predictor of Partial Control." *Linguistic Inquiry* 49: 379–393. https://doi.org/10.1162/LING_a_00276

Benmamoun, Elabbas, Archna Bhatia, and Maria Polinsky. 2009. "Closest Conjunct Agreement in Head Final Languages." *Linguistic Variation Yearbook* 9: 67–88. https://doi.org/10.1075/livy.9.02ben

Bowers, John. 2008. "On Reducing Control to Movement." *Syntax* 11: 125–143. https://doi.org/10.1111/j.1467-9612.2008.00103.x

Bruening, Benjamin. 2018. "Depictive Secondary Predicates and Small Clause Approaches to Argument Structure." *Linguistic Inquiry* 49: 537–559. https://doi.org/10.1162/ling_a_00281

Clark, Herbert, and Catherine Marshall. 1981. "Definite Reference and Mutual Knowledge." In *Elements of Discourse Understanding*, ed. by A. Joshi, B. Webber, and I. Sag, 10–63. Cambridge: Cambridge University Press.

Corbett, Greville. 1983. *Hierarchies, Targets and Controllers: Agreement Patterns in Slavic*. London: Croon Helm.

Creissels, Denis. 2011. "Impersonal Pronouns and Coreference: Two Case Studies." Unpublished ms., University of Lyon, France.

Duffey, Patrick. 2014. *Reclaiming Control as a Semantic and Pragmatic Phenomenon*. Amsterdam: John Benjamins.

Haug, Dag Trygve Truslew. 2014. "The Anaphoric Semantics of Partial Control." In *Proceedings of SALT* 24, ed. by T. Snider, S. D'Antonio, and M. Weigand, 213–233. Ithaca, NY: Cornell Linguistics Circle.

Hornstein, Norbert. 2001. *Move! A Minimalist Theory of Construal*. Oxford: Blackwell.

Irimia, Monica-Alexandrina. 2005. "Types of Secondary Predication." *Toronto Working Papers in Linguistics* 25: 20–29.

Kathol, Andreas. 2000. "Agreement and the Syntax-Morphology Interface in HPSG." In *Studies in Contemporary Phrase Structure Grammar*, ed. by R. Levine, and G. Green, 223–274. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511554421.007

Landau, Idan. 2000. *Elements of Control: Structure and Meaning in Infinitival Constructions*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-011-3943-4

Landau, Idan. 2016a. "Against the Null Comitative Analysis of Partial Control." *Linguistic Inquiry* 47: 572–580. https://doi.org/10.1162/LING_a_00222

Landau, Idan. 2016b. "Agreement at PF: An Argument from Partial Control." *Syntax* 19: 79–109. https://doi.org/10.1111/synt.12118

Lawler, John. 1972. *A Problem in Participatory Democracy*. Bloomington, IN: Indiana University Linguistics Club.

Marušič, Franc, Tatjana Marvin, and Rok Žaucer. 2003. "Depictive Secondary Predication with No PRO." Paper presented at the 5th European Conference on Formal Description of Slavic Languages, Leipzig University. sabotin.ung.si/~rzaucer_2008_Depictives-with-no-PRO.pdf

Morgan, Jerry. 1984. "Some Problems of Determination in English Number Agreement." In *Proceedings of the 1st Eastern States Conference in Linguistics*, ed. by G. Alvarez, B. Brodie, and T. McCoy, 69–78. Columbus, OH: The Ohio State University.

Nunes, Jairo. 2001. "Sideward Movement." *Linguistic Inquiry* 32: 303–344. https://doi.org/10.1162/00243890152001780

Nunes, Jairo. 2004. *Linearization of Chains and Sideward Movement*. Cambridge: MIT Press. https://doi.org/10.7551/mitpress/4241.001.0001

Pearson, Hazel. 2013. *The Sense of Self: Topics in the Semantics of de se expressions*. Boston, MA: Harvard University dissertation.

Pylkkänen, Liina. 2008. *Introducing Arguments*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262162548.001.0001

Ritchie, Katherine. 2014. "Collective Nouns: Polysemy and Predication." Paper presented at the 2014 Meeting of the Philosophical Linguistics and Linguistical Philosophy Conference in Tarrytown, NY.

Roberts, Ian. 1988. "Predicative APs." *Linguistic Inquiry* 19: 703–710.

Smith, Peter. 2015. "Feature Mismatches: Consequences for Syntax, Morphology and Semantics." University of Connecticut dissertation, Storrs, CT.

Smith, Peter. 2017. "The Syntax of Semantic Agreement in English." *Journal of Linguistics* 53: 823–863. https://doi.org/10.1017/S0022226716000360

Stowell, Tim. 1981. Origins of Phrase Structure. Los Angeles, CA: UCLA dissertation.

Wurmbrand, Susi. 2017. "Formal and Semantic Agreement in Syntax: A Dual Feature Approach." In *Language Use and Linguistic Structure*, ed. by J. Emonds, and M. Janebová, 19–36. Olomouc: Palacky University.

Person/Number exponents in imperative-enclitic contexts

Teresa Cabré & Maria Ohannesian
Universitat Autònoma de Barcelona

The goal of this paper is to analyse the partial or total deletion of the Person/Number exponents in imperative-enclitic contexts in Catalan and Spanish. Both languages lose phonological material in a specific context, but Spanish deletes the right-edge segment from the verb whereas Barcelona Catalan loses the left-edge segment of the clitic. Our aim is to account for the contrasting solutions shown by these two neighbouring languages using the same constraints but ranked differently in the Optimality Theory framework. OT allows us to put together different types of grammatical requirements, including distinctness requirements, particularly a positional OCP constraint on Person/Number features, and faithfulness conditions, specifically the condition related with morpheme realization, all of which interact with alignment and markedness constraints.

Keywords: P/N exponent, enclisis, reflexive, markedness, deixis

1. Introduction

It is well known that imperative forms tend to exhibit minimal or null morphology in natural languages. Sadock and Zwicky (1985: 72–73) analyse the presence of affixes in imperative forms in thirty-two languages of different typologies and report that "over half of the languages surveyed, in fact, employ an entirely affixless verbal base to indicate requests". For these authors, the relative frequency of the lack of affixes is a reflex of the inherent semantics of the imperative. "It is notionally future, so tense contrasts are unlikely. The subject naturally refers to the addressee, so second person inflection of the verb becomes redundant. The archetypical request is to do some task to completion, so indications of aspect are somewhat redundant. Semantic considerations cannot explain the suppression of conjugation-class distinctions, of course, but it remains a fact that this also occurs." The lack of person/tense affixation – especially in 2SG – follows from the semantics of imperative mood. Conversely, the lack of theme vowel has no semantic explanation because it is a verbal class marker, without any grammatical meaning.

https://doi.org/10.1075/rllt.16.03cab

True imperatives are directed at an addressee (the co-participant with the speaker in a speech act) and must be used in the present tense. They share these features with vocatives. As has been noted in the literature (Zanuttini 2008, among many others), vocatives and imperatives are inherently deictic, and this explains why exponence of second person, tense and aspect becomes redundant and is often lacking in imperatives. As verbal forms, imperatives are generated in inflexion, but they have to legitimate deictic features at the left periphery by means of intonation. As D'Alessandro and Oostendorp (2013: 11) state, "Core argumental functions can only be expressed by means of morphological and phonological tools, while prosody is reserved to the speech act." Therefore, those elements related to the speech act are expressed by means of prosody, that is, the corresponding intonational contour is the default exponent of the speech act features.

In this paper we focus on the phonological outcomes of Barcelona Catalan and Peninsular Spanish imperatives with reflexive enclitics, leaving outside our analysis the intonational contours in which the imperatives are uttered. Specifically, we account for the segmental deletion from adjacent Person/Number (henceforth P/N) exponents in imperative-enclitic contexts and we claim that the P/N feature identity in adjacent morphemes is responsible for this elision. The paper is structured as follows. In Section 2 we present the data and the morphological structure of imperatives and reflexive enclitic pronouns in these two languages. In Section 3 we define the OT constraints that are relevant for the phonological behaviour seen in this context. These constraints allow us to show in Section 4 how small changes in the ranking can account for the contrasting outcomes that the two languages display.

2. Morphological structure of imperatives and reflexive clitics

2.1 Imperatives

Deictic properties make the grammatical second person the true imperative form (Wurff 2007) and therefore, unlike what we see in all the other verbal forms, the second person imperative is the unmarked person. The confluence of P/N markedness in imperative 2SG explains the lack of phonological representation of this morpheme in Catalan and Spanish, whereas the 2PL always has a phonological exponent. For the other grammatical persons and negative forms, imperative resorts to subjunctive. The first person plural can also be used as an imperative because it includes the speaker and the addressee. The third person singular and plural may be used to express an imperative meaning only in a formal style. The strict morpheme order in the verbal paradigm is as follows: R (root), ThV (theme vowel) T/M (tense-aspect-mode), P/N (person-number). As is shown in (1) for

Catalan and (2) for Spanish, the P/N morpheme is realized phonologically only in the plural forms. The morphemes without phonological realization are represented by Ø. This morphological analysis is widely accepted. Note that the final vowel in first conjugation 2SG verbs is analysed as a theme vowel in Spanish (following Harris 1987) but as a tense marker in Catalan (as in Mascaró 1986) due to the different vowel systems in the two languages.

(1) Cat: rentar 'to clean'; bullir 'to boil'
   a. 2SG Imp: rent-Ø-a-Ø ['rentə]; bull-Ø-Ø-Ø ['buʎ]
   b. 2PL Imp: rent-e-Ø-u [rən'tɛw]; bull-i-Ø-u [bu'ʎiw]
   c. 1PL Imp: rent-e-Ø-m [rən'tɛm]; bull-i-Ø-m [bu'ʎim]

(2) Sp: lavar 'to clean'; decir 'to say'
   a. 2SG Imp: lav-a-Ø-Ø ['laβa]; di-Ø-Ø-Ø ['di]
   b. 2PL Imp: lav-a-Ø-d [la'βaˈð]; dec-i-Ø-d [de'θiˈð]
   c. 1PL Imp: lav-e-Ø-mos [la'βemos]; dig-a-Ø-mos [di'ɣamos]

The pronunciation of final *-d* as a weakened approximant in the Spanish 2PL form is quite widespread, due to the phonological markedness of /*d*/ in this position. On the other hand, in colloquial Peninsular Spanish, this form has been almost completely replaced by the infinitive form (e.g., *¡Lavad!* → *¡Lavar!* 'Clean!' you-PL, *¡Decid!* → *¡Decir!* 'Say!' you-PL). The infinitive is also used in negative forms, as in *¡No decir eso!* 'Don't say that!' you-PL), which alternates with the subjunctive (*¡No digáis eso!*).

2.2 Reflexive enclitics

Both Spanish and Catalan display proclisis in tensed verbal forms (Sp: *lo vi,* Cat: *el vaig veure* 'I saw it') and enclisis in imperatives, infinitives and gerunds (Sp: *voy a verlo,* Cat: *vaig a veure'l* 'I am going to see it').[1] In contrast with proclisis, enclisis exhibits more lexical cohesion with the verb in inflexional languages due to the adjacency or nearness of grammatical morphemes (e.g., the well-known phenomenon of mesoclisis in American Spanish: *siéntense* > *siéntensen* 'sit down' you-3PL.).

The phonological form of reflexive clitic pronouns in Catalan and Spanish are those represented in (3) and (4). Unlike Spanish, Catalan clitics display different phonetic forms depending on the context or even allomorphy in some varieties.

1. Not all Romance languages exhibit the same distribution of proclisis/enclisis. In European Portuguese, enclisis is possible in tensed forms (*Tu sentaste-**te** aqui* 'You sat down here'). It is worth noting that in literary language, Spanish also admits enclisis in tensed verbal forms (*Díjole* 'S/he told him/her).

(3)	Catalan	SG	PL
a.	1p	/m/	/nz/
b.	2p	/t/	/wz/
c.	3p	/s/	/s/

(4)	Spanish	SG	PL
a.	1p	/me/	/nos/
b.	2p	/te/	/os/
c.	3p	/se/	/se/

Following Bonet and Lloret (2005), but in opposition to Mascaró (1986), we assume that /nz/ and /wz/ in Catalan are unanalysable phonological forms and therefore do not have internal morphological structure. And contrary to Harris and Halle (2005), we extend this analysis to the Spanish clitics /nos/ and /os/ because the features they represent are essentially the same as those in the verbal morphemes. Although these final sibilants /z/ and /s/ coincide with the phonological form of plural in nominals and other clitic forms, their behaviour in enclitic position compels us to analyse these clitics as morphological units, as we will see below. Unlike other clitic pronouns in Barcelona Catalan, first and second person plural have the same form in proclitic as in enclitic position (e.g., *[ənz] van cridar* / *van crid[ˈanzə]* 'they called to us'; *[uz] van renyar / van reny[ˈawzə]* 'they told you-PL off'. In addition, we also assume that the high vocoid of the second person plural is a glide due to its phonological behaviour in non-initial position.[2]

2.3 Imperative-enclitic context resolutions

When the imperative plural forms are used with enclitic reflexives, a segmental deletion is triggered in both languages: Barcelona Catalan (5) loses material from the clitic whereas Spanish (6) deletes material from the verbal form.

(5) a. renteu#us 'clean yourselves' → [rənˈtɛwzə]
b. rentem#ns 'clean ourselves' → [rənˈtɛmzə]

(6) a. lavad#os 'clean yourselves' → lavaos
b. lavemos#nos 'clean ourselves' → lavémonos

In other words, in both languages whenever the exponence of identical values for P/N features from the imperative and enclitic reflexive stand adjacent, they

2. The impersonal/third person clitic /s/ is the sole clitic pronoun that can precede the first and second person in Catalan: *Se'ns obren les portes* [sənz] 'Doors are opening for us' / *Se us obren les portes* [səwz] 'Doors are opening for you-PL'. Compare with: *S'ho troben* [su] 'They find it'.

come into collision, and this forces a segment to be cancelled out. It could be said that both languages avoid deleting the whole P/N exponent in compliance with the universal principle of structural preservation. This explains why Catalan deletes the first segment from the clitic and not the monosegmental P/N morph from the verb. But the behaviour seen in Spanish is quite baffling given that the P/N morpheme from the verb is always the target of deletion, even when the whole P/N morpheme /d/ is removed. Conversely, the use of the morphological third person to address someone formally does not trigger any phonological variation in either language (e.g., Sp: *Lávense/lávese la ropa*; Cat: *Rentin-se/renti's la roba* 'Clean the clothes yourselves/yourself') due to the fact that the deictic person and the exponent of the grammatical person do not coincide. It should be pointed out that clitics do not trigger deletion in non-reflexive contexts: Sp.: *lavémoslas, *lavémolas* 'let us clean them-f'; Cat.: *rentem-les, *rentem-es*.

It is also necessary to say that the Spanish obligatory deletion in sentences like *lavaos* 'clean yourselves' should not be confused with the deletion of /*d*/ in participle forms, a very common – but not obligatory – phenomenon in many varieties, as in *lavado* → [laˈβaŏ] 'washed'.

As we noted above, Peninsular Spanish usually takes the infinitive form as the imperative second person plural. This has been analysed as a lenition of the final morpheme /d/ (Alcoba 1999; RAE 2009). However, we claim that this is the true infinitive form used as the imperative second person plural, which has the advantage of not triggering conflict with any adjacent features (e.g., *lavaros* 'clean yourselves').

It is fair to say that deletion in imperative-enclitics has been described and analysed in the literature for both Catalan and Spanish separately (among others, Fabra 1913–14; Bonet & Lloret 2005 for Barcelona Catalan and Alcoba 1999; RAE 2009 for Spanish). Our contribution is to compare deletion in the two languages and show that the opposite paths they take in deleting elements are in fact subject to the very same constraints.

It is also important to point out that the second person singular enclitic /t/ in Catalan seems to show phonological behaviour different from that shown by the first and second plural /nz/ and /wz/, since the former is realized without vowel epenthesis (e.g., *renta't*) whereas the latter add a final vowel which is not needed for standard syllabification (*rentem*-[zə], *renteu*-[zə]). Indeed, hypothetical forms such as **rentem's* or **renteu's* (spelled with non-standard orthography) should be no less acceptable than words like *enciams* 'lettuces' or *peus* 'feet'. The problem seems to lie in the morphological structure of the syllabic coda: a complex coda cannot be filled by a verbal element and a clitic element (Bonet & Lloret 2005).

3. OT account

Our goal is to account for the erasure of segments that compose (part of) the exponents of the P/N morpheme in imperative-enclitic contexts. The agreement of P/N morphosyntactic features between the verbal ending and the adjacent reflexive clitic motivates that loss, but with two opposite solutions: while Catalan deletes the left edge segment of the clitic, Spanish deletes the right edge segment of the verbal ending. The verbal P/N exponent is always monosegmental in Catalan (*cante-m* 'sing-1PL'; *cante-u* 'sing-2PL'), as it is also in the second person plural in Spanish (*canta-d* 'sing-2PL'). However, as noted above, multisegmental P/N exponents (*cante-mos* 'sing-1PL' in Spanish) cannot be split up into two morphemes even when the last one exhibits a phonological coincidence with the plural morph, since there is no parallel between the singular-plural exponents of the corresponding persons. Optimality Theory (Prince & Smolensky 1993/2004) allows us to consider requirements of different grammatical natures in the same ranking and show how phonology and morphology interact with each other.

In order to explain what prevents two adjacent morphemes with the same grammatical features – in this case, the P/N exponent of the imperative and the reflexive pronoun – from being phonetically realized, we propose the negative constraint *αP/N-αP/N, which should be understood as the morphological translation of the phonological Obligatory Contour Principle (McCarthy 1986, among others).[3] This constraint defined in (7) must be interpreted as it forces the deletion of one edge-segment in adjacent morphemes with identical P/N values.

(7) *αP/N-αP/N: Adjacency of complete morphemes with identical P/N values are banned.

The vanishing of the whole P/N exponent /-d/ in Spanish constitutes a fatal violation of the universal principle of Structure Preservation (Kiparsky 1985). In the OT framework, this phonological principle has been adapted as a faithfulness constraint named Realize-Morpheme (RM) (Oostendorp 2005), which is defined in (8). RM is the counterweight to *αP/N-αP/N and thereby essential to prevent cancellation of the whole exponent.

3. From the Minimalist syntactic perspective, Richards (2010) posits the Distinctiveness condition, a syntactic condition that blocks two adjacent elements with identical features in the same domain.

(8) Realize-Morpheme (RM): For every morpheme in the input, some phonological element should be present in the output. (Oostendorp 2005)

As Oostendorp (2005: 118) notes, the central notion that allows RM to be interpreted correctly is 'recoverability', that is, the possible information lost in some process must be recoverable from the context, that context being in the case under study the adjacent exponent of P/N.

Notice that RM blocks total deletions of input morpheme. However, like any constraint, it can be violated. In this case *αP/N-αP/N, which entails the deletion of the morpheme /d/, dominates RM. The recoverability of P/N features is guaranteed by the adjacent morpheme with the same P/N features.

Given that the deletion cases of interest here take place in the imperative-enclitic boundary, it is necessary to resort to an alignment constraint in order to formally express the context where the process occurs. The constraint AlignV/Cl (Bonet & Lloret 2005) defined in (9) aligns the right edge of the verb with the left edge of the clitic, thus preventing the irruption of space between one edge and the other.

(9) AlignV/Cl: Align the right edge of the verb with the left edge of the pronominal clitic. (Bonet & Lloret 2005)

Finally, the faithfulness constraint Max (McCarthy 1995) preserves from deletion the phonological material of the input. It is defined in (10).

(10) Max: Every element in the input has a correspondent in the output.

The tableaux in (11) and (12) display how the hierarchy of the above constraints correctly accounts for the output forms in (11a) and (12a) because they only violate Max, which is dominated by all the other constraints. None of the other three constraints (*αP/N-αP/N, Realize-Morpheme, AlignV/Cl) dominates any of the others. Nonetheless, this hierarchy does not take into account the analysis of the final epenthetic vowel which is subject to syllabic constraints, as we noted in Section 2, so the constraint responsible for preventing the insertion of epenthetic elements (Dep) is not included. The hash mark (#) indicates the verb-clitic boundary; we use it to make clear which phonological process has occurred in each candidate. Candidates (11b) and (12b) have introduced an epenthetic vowel in the verb-clitic boundary, thus disobeying the alignment constraint. Candidates (11c) and (12c) delete the whole exponent of the P/N morpheme, fatally violating Realize-Morpheme. Candidates (11d) and (12d) violate *αP/N-αP/N but they would also be ruled out by impossible syllabification.

(11) Cat: V-1PL # cl-1PL

rəntɛ-m#nz	*αP/N-αP/N	AlignV/Cl	RM	Max
a. ☛rəntɛm#zə				*
b. rəntɛm#ənz		*!		
c. rəntɛ#nzə			*!	*
d. rəntɛm#nzə	*!			

(12) Cat: V-2PL # cl-2PL

rəntɛ-w#wz	*αP/N-αP/N	AlignV/Cl	RM	Max
a. ☛rəntɛw#zə				*
b. rəntɛw#əwzə		*!		
c. rəntɛ#wzə			*!	*
d. rəntɛw#wzə	*!			

The tableau in (12) requires some additional comments because it deals with a sequence of two identical high rounded glides. Perhaps its erasure or fusion could be accounted for phonologically by means of the general conditions that determine vowel contact resolution in Catalan (for instance, compare it with the sentence *Beveu wisky* [bəβɛwíski] 'Drink whisky'). However, the constraint ranking proposed for the first person plural fits perfectly with the second person plural and allows us to see that what we are dealing with here is a deletion rather than a fusion, and the deleted glide belongs to the clitic, not to the verb.

Let us now turn to the Spanish outcomes. The main difference from Catalan is that the deletion always takes place on the verb, the clitic exponents remaining intact. The systematic preservation of the whole clitic compels us to resort to a specific anti-deletion constraint on clitics (Max_{Cl}), defined in (13). Max_{Cl} must dominate Realize-Morpheme to force the vanishing of /d/ from the second person plural verbal form.

(13) Max_{CL}: Every clitic in the input has a correspondent in the output.

The tableau in (14) shows how candidate (14a) is the only possible output. The Spanish equivalent form of the Catalan winner (14d) fatally violates Max_{Cl}. Otherwise, however, this candidate would be ruled out because the left-edge segment

deletion of the clitic would entail a complete homophony with the second person plural clitic. The remaining candidates also crucially disobey one or another high-ranked constraint.

(14) Sp: V-1PL # cl-1PL

lave-mos#nos	*αP/N-αP/N	ALIGNV/CL	MAX_{CL}	RM	MAX
a. ☛lavémo#nos					*
b. lavémos#nos	*!				
c. lavémose#nos		*!			*
d. lavémos#os			*!		

Tableau (15) displays how the constraint ranking acts on imperative-enclitic second person plural contexts. As is well known, a full form such as *lavad#os* is completely ruled out even though its syllabification would be optimal. This means that markedness constraints related to the syllable structure (such as Onset or *Coda) are not relevant for the cases under study and are therefore not included in the constraint ranking. Candidate (15a) wins even though it violates the RM constraint, which is dominated by *αP/N-αP/N, ALIGNV/CL and MAX_{CL} in Spanish. The remaining candidates are ruled out because they all violate one of the dominant constraints.

(15) Sp: V-2PL # cl-2PL

lava-d#os	*αP/N-αP/N	ALIGNV/CL	MAX_{CL}	RM	MAX
a. ☛lava#os				*	*
b. lavad#os	*!				
c. lavade#os		*!			*
d. lavad#s			*!		

Let us return now to the fact noted above that Peninsular Spanish usually takes the infinitive form as the second person plural imperative. Three factors seem to motivate this preference: the neutral character of the infinitive as a non-finite verbal form, the strong phonological markedness of the final -d, which is pronounced as a weakened approximant in most Peninsular varieties, and morphological pressure to distinguish the second person plural from the singular. Since a non-finite verbal form has no P/N exponent, the clash of adjacent exponents

with identical P/N values cannot occur, and the *αP/N-αP/N constraint therefore becomes superfluous. The tableau in (16) displays how the same constraint ranking accounts for the correct output.

(16) Sp: V- Inf. # cl-2PL

lava-r#os	*αP/N-αP/N	ALIGNV/CL	MAX_{CL}	RM	MAX
a. ☛lavar#os					
b. lava#os				*!	*
c. lavare#os		*!			
d. lavar#s			*!		*

To sum up, the ranking of constraints in Catalan and Spanish are shown in (17) and (18). Note that the crucial difference between the two languages lies in the hierarchical order of the constraints REALIZE-MORPHEME and MAX_{CL}, since RM is high-ranked in Catalan but dominated in Spanish, while on the contrary MAX_{CL} is demoted to the bottom of the ranking because it is not active in Catalan but high-ranked in Spanish.

(17) Cat: *αP/N-αP/N, ALIGNV/CL, REALIZE-MORPHEME (RM) >> MAX, MAX_{CL}

(18) Sp: *αP/N-αP/N, ALIGNV/CL, MAX_{CL} >> REALIZE-MORPHEME (RM) >> MAX

4. Conclusions

Deixis and markedness are responsible for the phonological lack of second person singular morphemes in imperatives and prosody is the true phonological exponent for modality. Tense does not need phonological realization because it can only have one interpretation. Therefore it seems clear that the obligatory segmental deletion in the varieties under study is triggered by a morphological need, to wit, the need to avoid the entire phonetic realization of two adjacent morphemes with the same Person/Number features.

Avoidance of identical adjacent features is frequent in all grammar systems, an issue that has been analysed from the perspectives of many different approaches such as *Clash (Kager 1994), Morphological Haplology (de Lacy 1999), OCP (McCarthy 1986), syntactic OCP (Yip 1998) or the Distinctness Condition (Richards 2010). Our OT approach allows us to combine phonological and morphological aspects in order to account for the deletion that occurs in imperative-enclitic contexts in both Spanish and Catalan.

Funding

Spanish Ministerio de Economía, Industria y Competitividad FFI2016-76245-C3-1-P AGAUR-Generalitat de Catalunya 2017 SGR 634

Acknowledgment

We would like to thank participants at the 47th Linguistic Symposium on Romance Languages held in April 2017 at the University of Delaware, as well as an anonymous reviewer who provided invaluable comments. This work has been supported by grants FFI2016-76245-C3-1-P (AEI/FEDER, UE) from the Spanish Ministerio de Economía, Industria y Competitividad and 2017 SGR 634 from AGAUR-Generalitat de Catalunya.

References

Alcoba, Santiago. 1999. "La Flexión Verbal." In *Gramática descriptiva de la lengua española*, ed. by I. Bosque, and V. Demonte, 4916–4991. Madrid: Espasa.

Bonet, Eulàlia, and Maria-Rosa Lloret. 2005. "More on Alignment as an Alternative to Domains: The Syllabification of Catalan Clitics." *PROBUS* 17: 37–78. https://doi.org/10.1515/prbs.2005.17.1.37

D'Alessandro, Roberta, and Marc van Oostendorp. 2013. "Phi Features for Phonology; Edge Features for Prosody. Insights into the Syntax-PF interface." Ms., Leiden University.

de Lacy, Paul. 1999. "Morphological Haplology and Correspondence." In *University of Massachusetts Occasional Papers in Linguistics* 25, ed. by P. de Lacy, and A. Nowak, 51–88. Amherst, MA: Graduate Linguistics Students Association.

Fabra, Pompeu. 1913–1914. "Els mots àtons en el parlar de Barcelona." *Butlletí de Dialectologia Catalana*, I (1913): 7–17; II (1914): 1–6.

Harris, James. 1987. "The Accentual Patterns of Verb Paradigm in Spanish." *Natural Language and Linguistic Theory* 5: 61–90. https://doi.org/10.1007/BF00161868

Harris, James, and Morris Halle. 2005. "Unexpected Plural Inflections in Spanish: Reduplication and Metathesis". *Linguistic Inquiry* 36: 195–222. https://doi.org/10.1162/0024389053710710

Kager, Réné. 1994. "Ternary Rhythm in Alignment Theory." Ms., Utrecht University.

Kiparsky, Paul. 1985. "Some Consequences of Lexical Phonology". *Phonology Yearbook*, 85–138. Cambridge: Cambridge University Press.

Mascaró, Joan. 1986. *Morfologia*. Barcelona: Enciclopèdia Catalana.

McCarthy, John. 1986. "OCP Effects: Gemination and Antigemination." *Linguistic Inquiry* 17: 207–263.

McCarthy, John. 1995. "Extensions of Faithfulness: Rotuman Revisited." Ms., UMass-Amherst.

van Oostendorp, Marc. 2005. "Expressing Inflection Tonally." *Catalan Journal of Linguistics* 4: 107–126. https://doi.org/10.5565/rev/catjl.115

Prince, Alan, and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA, & Oxford: Blackwell.

Real Academia Española. 2009. "La flexión verbal." In *Nueva gramática de la lengua española*, 181–336. Madrid: Espasa.

Richards, Norvin. 2010. *Uttering Trees* (LI Monograph 56). Cambridge: MIT Press. https://doi.org/10.7551/mitpress/9780262013765.001.0001

Sadock, Jerrold, and Arnold Zwicky. 1985. "Speech Act Distinctions in Syntax." In *Language Typology and Syntactic Structure, Vol. 1: Clause Structure*, ed. by T. Shopen, 155–196. Cambridge: Cambridge University Press.

van der Wurff, Wim. 2007. "Introduction." *Imperative Clauses in Generative Grammar*, 1–94. Amsterdam: John Benjamins. https://doi.org/10.1075/la.103

Yip, Moira. 1998. "Identity Avoidance in Phonology and Morphology." In *Morphology and its Relation to Phonology and Syntax*, ed. by S. Lapointe, D. Brentari, and P. Farrel, 216–246. Stanford: CLSI Publication.

Zanuttini, Raffaella. 2008. "Encoding the Addressee in the Syntax. Evidence from English Imperative Subjects." *Natural Language and Linguistic Theory* 26: 185–218. https://doi.org/10.1007/s11049-007-9029-6

French causal *puisque*-clauses in the light of (not)-at-issueness

Isabelle Charnavel
Harvard University

This paper examines the discourse status of French causal clauses introduced by *puisque* 'since'. *Puisque*-clauses are associated with two implications: the relation expressed by *puisque* and the content of their clause. Several diagnostics show that neither implication is at-issue and that the two implications belong to two different types of projective content (cf. Tonhauser et al. 2013). This is due to syntactic and lexical reasons: the relation expressed by *puisque* is not at-issue because *puisque*-clauses modify high (evidential and speech act) phrases that are not at-issue; the content of the *puisque*-clause is not at-issue because *puisque* is lexically factive, i.e. selects true facts. *Puisque* thus exemplifies an unrecognized type of double presuppositional trigger.

Keywords: syntax/semantics, causal clauses, (not)-at-issueness, presupposition, projection, evidentiality, speech act, French

Introduction

A wide range of constructions has been shown to be associated with backgrounded implications that tend to survive even when their trigger is embedded under operators that usually block the implications in their scope (i.e. they project, see Tonhauser et al. 2013, a.o.). This includes all implications standardly analyzed as presuppositions or as conventional implicatures, which are triggered by various types of words, constructions or utterances (e.g. factive verbs or appositives, among many others). This article focuses on one type of trigger that remains understudied, namely (some) adjunct clauses, by studying the particular case of French causal clauses introduced by the subordinator *puisque* 'since' (*puisque*-clauses, henceforth) as illustrated in (1)–(2).

(1) *Lise est chez elle, puisque la lumière est allumée*
'Liz is home, since the light is on'

(2) *Allons prendre un verre, puisque tu insistes*
'Let's go for a drink, since you insist'

https://doi.org/10.1075/rllt.16.04cha

Puisque-clauses are associated with two implications: the relation expressed by *puisque* (e.g. the evidential relation between the light and Liz's presence in (1)) and the content of the *puisque*-clause (the fact that the light is on in (1)). The goal of this paper is to show that the two implications qualify as two types of projective, not-at-issue content, and to derive their not-at-issueness from two different sources (syntactic and lexical). *Puisque*-clauses thus promise to shed new light on the properties of double triggers and the constraints on their implications.

The empirical part of this paper will consist in demonstrating, based on specific diagnostics, that neither implication associated with *puisque*-clauses is at-issue (Section 1) and that each implication falls into a different type of projective content (Section 2). The analytical part of the paper (Section 3) will attribute a syntactic source to the not-at-issueness of one implication (the relation expressed by *puisque*) and a lexical one to the not-at-issueness of the other (the content of the *puisque*-clause).

1. Not-at-issueness of *puisque*-clauses

Unlike clauses introduced by *parce que* 'because' that usually[1] describe the cause of the matrix event as in (3) below, *puisque*-clauses provide some evidence for (believing) the matrix proposition as in (1) or (4) below, or some reason for the matrix speech act as in (2) or (5) (cf. Groupe λ-1 1975; Ducrot 1983, a.o.).[2] In both cases, they trigger two types of implications: the content of the subordinate clause (call it the B-implication) as specified in (4a) and (5a), and the relation expressed by *puisque* (call it the *puisque*-implication) as specified in (4b) and (5b).

---

1. At least for some speakers, clauses introduced by *parce que* 'because' can also (with a different prosody involving an intonational break between the matrix and the adjunct clause) have the same semantic contribution as *puisque*-clauses in (1)–(2). Cf. Rutherford 1970; Groupe λ-1 1975; Sæbø 1991, a.o.

2. That's not to say that the following sentence is unacceptable:

   (i) *Lise est partie puisqu'elle était fatiguée*
   'Liz left since she was tired'

But (i) does not have the same meaning as (3) and cannot therefore be used in the same contexts: in (3), the *parce que*-clause provides the cause of Liz's departure; but in (i), the *puisque*-clause provides a justification for my saying that Liz left; moreover, Liz's tiredness is presented as the main point of the utterance in (3), unlike in (i). These differences are revealed by the different results obtained in the tests for not-at-issueness presented in this section and in the next section. Note that this is in the spirit of Groupe λ-1 (1975) that analyzes *puisque* (unlike *parce que*) as a connector linking two different speech acts.

(3) *Lise est partie parce qu'elle était fatiguée*
'Liz left because she was tired'

(4) *Lise est partie, puisque son manteau n'est pas là*
'Liz left, since her coat is not here'
a. Liz's coat is not here [B-implication]
b. the absence of her coat indicates that Liz left [*puisque*-implication]

(5) *Lise est partie, puisque tu veux tout savoir*
'Liz left, since you must know everything'
a. the addressee must know everything [B-implication]
b. the addressee's curiosity is a reason for the speaker to inform her that Liz left [*puisque*-implication]

In both types of *puisque*-clauses, neither implication is central to the discussion: several diagnostics (detailed below) show that they are not-at-issue, which makes *puisque*-clauses similar to other types of not-at-issue content like appositives or factive clauses.

1.1 Answer to question under discussion

The first diagnostic relies on the ability of the target implication to address the Question Under Discussion (QUD, see Roberts 1996): at-issue content can address the QUD, while not-at-issue cannot. Indeed, a proposition is said to be at-issue if it is relevant to the QUD and this can be recognized by the addressee (Simons et al. 2010, a.o.).

Applying this diagnostic to *puisque*-clauses, the infelicitous dialogs in (6a) and (7a) show that the content of the *puisque*-clause (the B-implication) is not-at-issue, unlike the content of the matrix clause (call it the A-implication, see (6c) and (7c)). The relation expressed by *puisque* (the *puisque*-implication) is not at-issue either, as evidenced by (6b) and (7b): unlike clauses introduced by *parce que* in (8), *puisque*-clauses cannot be used to answer *why*-questions – or any other type of question.

(6) a. **Q:** *Où est le manteau de Lise?* [B-implication]
'Where is Liz's coat'
A: *#Lise est partie, puisque son manteau n'est pas là*
'#Liz left, since her coat is not here'

b. **Q:** *Pourquoi est-ce que Lise est partie?/Comment tu sais que Lise est partie?* [*puisque*-implication]
'Why did Liz leave? / How do you know that Liz left?'
A: *#Lise est partie, puisque son manteau n'est pas là*
'#Liz left, since her coat is not here'

c. **Q:** *Où est Lise?* [A-implication]
'Where is Liz?'

A: *Lise est partie, puisque son manteau n'est pas là*
'Liz left, since her coat is not here'

(7) a. **Q:** *Qu'est-ce que je veux savoir?* [B-implication]
'What should I know?'

A: *#Lise est partie, puisque tu veux tout savoir*
'#Liz left, since you must know everything'

b. **Q:** *Pourquoi est-ce que Lise est partie?/Pourquoi devrais-tu me dire où est Lise?* [*puisque*-implication]
'Why did Liz leave?/Why should you tell me where Liz is?'

A: *#Lise est partie, puisque tu veux tout savoir*
'#Liz left, since you must know everything'

c. **Q:** *Où est Lise?* [A-implication]
'Where is Liz?'

A: *Lise est partie, puisque tu veux tout savoir*
'Liz left, since you must know everything'

(8) **Q:** *Pourquoi est-ce que Lise est partie?* [*parce que*-implication]
'Why did Liz leave?'

A: *Lise est partie parce qu'elle était fatiguée*
'Liz left because she was tired'

1.2 Challengeability

The second diagnostic is based on the fact that only at-issue content can be directly accepted or rejected by an addressee (Tonhauser 2012, a.o.); not-at-issue content, however, can only indirectly be challenged with utterances such as "Hey, wait a minute!" (see von Fintel 2004, a.o.). One way to diagnose (not)-at-issueness is thus to explore the acceptability of utterances where an assent is followed by an adversative utterance that conveys a hypothesized at-issue or not-at-issue content (Tonhauser's 2012 diagnostic #1c). The assumption behind the diagnostic is that such utterances are contradictory, therefore unacceptable, when the continuations convey at-issue content, but acceptable when they convey not-at-issue content. The diagnostic is applied to *puisque*-clauses below.

(9) *Lise est partie, puisque son manteau n'est pas là*
'Liz left, since her coat is not here'

a. *Oui, c'est vrai, mais son manteau est là, en fait*
'Yes, true, but her coat is in fact here' [B-implication]

b. *Oui, c'est vrai, mais l'absence de son manteau ne prouve rien : elle ne l'avait pas pris*
'Yes, true, but the absence of her coat does not prove anything: she didn't take it with her' [*puisque*-implication]

c. *#Oui, c'est vrai, mais Lise n'est pas partie en fait*
'#Yes, true, but Liz did not leave, in fact' [A-implication]

(10) *Lise est partie, puisque tu veux tout savoir*
'Liz left, since you must know everything'

a. *Oui, c'est vrai, mais je n'ai pas demandé à tout savoir*
'Yes, true, but I didn't ask to know everything' [B-implication]

b. *Oui, c'est vrai, mais ce n'est pas parce que je veux tout savoir que tu dois tout me dire*
'Yes, true, but it is not because I want to know everything that you need to tell me everything' [*puisque*-implication]

c. *#Oui, c'est vrai, mais Lise n'est pas partie en fait*
'#Yes, true, but Liz did not leave, in fact' [A-implication]

Both (a)–(b) responses in (9) and (10) consist of assents followed by adversative utterances that convey the content of the B-implication and that of the *puisque*-implication, respectively. Their acceptability confirms that the B-implication and the *puisque*-implication are not at-issue. Just as above, this contrasts with the status of the A-implication, as shown in responses (9c) and (10c), and with that of the *because*-implication, as shown in (11) below: these responses are contradictory, which confirms that the A-implication and the *parce que*-implication are at-issue.

(11) *Lise est partie parce qu'elle était fatiguée* [*parce que*-implication]
'Liz left because she was tired'
#Oui, c'est vrai, mais la fatigue n'était pas la cause de son départ
'#Yes, true, but her tiredness was not the cause of her departure'

1.3 Non-focusability

A further observation that corroborates this generalization is that *puisque*-clauses cannot be focused (cf. Groupe λ-1 1975; Sæbø 1991, a.o.). For instance, neither type of *puisque*-clause can be clefted, as shown in (12), while *parce que*-clauses can, as illustrated in (13).

(12) a. **C'est puisque son manteau n'est pas là que Lise est partie*
'*It is since her coat is not here that Liz left'

b. **C'est puisque tu veux tout savoir que Lise est partie*
'*It is since you must know everything that Liz left'

(13) *C'est parce qu'elle était fatiguée que Lise est partie*
'It is because she was tired that Liz left'

Similarly, (14) and (15) show that unlike *parce que*-clauses, *puisque*-clauses cannot be associated with a focus particle like *seulement* 'only'.

(14) a. *#Lise est seulement partie puisque son manteau n'est pas là*
'#Liz only left since her coat is not here'
[Intended: the only piece of evidence indicating that Liz left is the absence of her coat]

b. *#Lise est seulement partie puisque tu veux tout savoir*
#Liz only left since you must know everything.
[Intended: the only reason why I am asserting that Liz left is your willingness to know everything]

(15) *Lise est seulement partie parce qu'elle était fatiguée*
Liz only left because she was tired.
[i.e. the only reason why Liz left is her tiredness]

Given that not-at-issue content cannot usually be focused,[3] this supports the hypothesis that *puisque*-clauses are not at-issue.

2. Two types of projective content

2.1 Projection

It has been observed that all and only not-at-issue content (including both so-called presuppositions and conventional implicatures) project (Simons et al. 2010, Tonhauser et al. 2013, a.o.), i.e. survive under a negation, question, modal or conditional operator (cf. family-of-sentences test in Chierchia & McConnell-Ginet 1990). This correctly predicts that the two implications associated with *puisque*-clauses are not affected by these operators: in (16a-d), the speaker remains committed to the fact that Liz's coat is not here (B-implication) and that the absence of

3. The relation between not-at-issueness and focusability is more complex (see Abrusán 2016; Sæbø 2016, a.o.). In particular, the entire not-at-issue content cannot be focused, but part of it can contain contrastive focus as illustrated below.

(ii) **Q:** *Je ne vois pas le sac de Lise, est-ce qu'elle est partie?*
'I cannot see Liz's purse, has she left?'

A: *Non, elle n'est pas partie, puisque son manteau, lui, est toujours là*
'No, she hasn't left, since her COAT is still here'

Liz's coat is a piece of evidence indicating that she left (*puisque*-implication). This contrasts with *parce que*-clauses in (17), where the reason for Liz's departure (the *parce que*-implication) can be negated, questioned and modalized (cf. Groupe λ-1 1975, a.o.).

(16) a. *Ce n'est pas le cas que Lise est toujours là puisque son manteau n'est pas là*
'It is not the case that Liz is still here since her coat is not here'

b. *Est-ce que Lise est partie puisque son manteau n'est pas là?*
'Did Liz leave since her coat is not here?'

c. *Il est possible que Lise soit partie puisque son manteau n'est pas là*
'It is possible that Liz left since her coat is not here'

d. *Si Lise est partie puisque son manteau n'est pas là, je suis déçu*
'If Liz left since her coat isn't here, I am disappointed'

(17) a. *Ce n'est pas le cas que Lise est partie parce qu'elle était fatiguée*
'It is not the case that Liz left because she was tired'
[Intended: it is not Liz's tiredness that caused her to leave]

b. *Est-ce que Lise est partie parce qu'elle était fatiguée ?*
'Did Liz leave because she was tired?'
[Intended: was Liz's tiredness the reason why she left?]

c. *Il est possible que Lise soit partie parce qu'elle était fatiguée*
'It is possible that Liz left because she was tired'
[Intended: Liz's tiredness is a plausible reason why she left]

d. *Si Lise est partie parce qu'elle était fatiguée, je ne vais pas la déranger*
'If Liz left because she was tired, I will not bother her'
[Intended: if Liz's tiredness is the reason why she left, I will not bother her]

Both the B-implication and the *puisque*-implication thus fall into the class of projective content. The tests explained below furthermore show that they fall into two different subclasses of projective content.

2.2 Strong contextual felicity constraint

Projective contents differ with respect to the constraints they impose on the context: some utterances with projective content are only acceptable in contexts that entail or imply it (strong contextual felicity, see Tonhauser et al. 2013); others do not have this requirement (cf. presupposition accommodation). Example (18) shows that the B-implication falls into the latter case: the addressee Paul need not know that the neighbors' radio is on for Claire's utterance to be felicitous.

(18) *[Context: Claire and Paul's neighbors told them that they turn their radio on every time they leave to turn away potential thieves. Paul is taking a shower and cannot hear anything. Claire comes back from the garden where she heard their radio and tells him (very loudly so that he can hear):]* [B-implication]

Les voisins sont partis, puisque leur radio est allumée
'The neighbors have left, since their radio is on'

The *puisque*-implication, however, is subject to the strong contextual felicity constraint: Claire's utterance is infelicitous in the context of (19) where Paul is not aware of the evidential relation expressed by *puisque* (i.e. the radio indicates that the neighbors left).

(19) *[Context: When Paul was away, Claire and Paul's neighbors told Claire that they turn their radio on every time they leave to turn away potential thieves. Paul now hears the neighbors' loud radio and asks Claire to talk to them. She replies:]* [*puisque*-implication]

#Les voisins sont partis, puisque leur radio est allumée
'#The neighbors have left, since their radio is on'

2.3 Obligatory local effect

The two implications associated with *puisque* do not differ, however, with respect to the second diagnostic identified by Tonhauser et al. (2013) to distinguish between subclasses of projective content. This diagnostic detects projective content that must be in the scope of an attitude verb when its trigger is syntactically embedded in its complement (obligatory local effect). Both the *puisque*-implication and the B-implication are subject to this effect.

This can only be shown using the type of *puisque*-clauses such as (1) or (4) that provide evidence for the matrix proposition, because only this type can be embedded in attitude contexts. For instance, (4) can be transposed into an indirect discourse as in (20):[4] embedding of the *puisque*-clause under the attitude verb *pense* 'thinks' is here enforced syntactically by its fronting within the embedded clause and pragmatically by the use of the matrix negative verb *refuse* 'refuses'.

4. Just like epistemic modals (see Anand & Hacquard 2013), evidential *puisque*-clauses can only be embedded under predicates of acceptance like *croire* 'believe' or *penser* 'think', not under desideratives like *vouloir* 'want' or directives like *demander* 'demand', as shown by the contrast between (iiia) and (iiib).

(iii) a. *Paul pense que puisque son manteau n'est pas là, Lise est partie*
'Paul thinks that since her coat is not here, Liz left'

b. **Paul {veut/demande} que puisque son manteau n'est pas là, Lise parte*
'*Paul {wants/demands} that since her coat is not here, Liz leave'

(20) *Paul refuse de croire que puisque son manteau n'est pas là, Lise est partie*
'Paul refuses to believe that since her coat is not here, Liz left'

However, *puisque*-clauses like (2) or (5) that express a reason for the matrix speech act are not embeddable: unlike (4), (5) cannot be transposed into an indirect discourse, as shown by Example (21), where embedding of the *puisque*-clause is enforced by its fronting within the embedded clause and can be further guaranteed by the *de dicto* non *de re* reading of *pénible* 'annoying'; in that case, the *puisque*-clause cannot be interpreted as providing a reason for Paul's reported speech act. This can be accounted for by the usual unembeddability of speech acts (Krifka 2014, a.o.).

(21) *#Paul annonce que puisque son pénible interlocuteur veut tout savoir, Lise est partie*
'#Paul announces that since his annoying interlocutor must know everything, Liz left'

If we use embeddable *puisque*-clauses to test for the obligatory local effect of the implications associated with them, we observe that both the B-implication and the *puisque*-implication are subject to this effect: both (22) and (23)[5] are contradictory, which shows that the two implications must be interpreted in the scope of *pense* 'thinks'.

(22) *#Paul pense que puisque leur radio est allumée, les voisins sont partis et que leur radio n'est pas allumée*
'#Paul thinks that since their radio is on, the neighbors left and that their radio is off' [B-implication]

(23) *#Paul pense que puisque leur radio est allumée, les voisins sont partis et il ignore que les voisins allument leur radio quand ils partent*
'#Paul thinks that since their radio is on, the neighbors left and he does not know that the neighbors turn their radio on when they leave'
[*puisque*-implication]

Thus, the two implications associated with *puisque*-clauses belong to two different subclasses of projective content: the B-implication, which is not subject to the strong contextual felicity constraint, but has obligatory local effect, falls into Tonhauser et al.'s 2013 class C, just like the content of the complement of factive verbs; the *puisque*-implication, which has both the strong contextual felicity constraint

5. The test is applied slightly differently in (22) and (23) (the continuations start with *and [he thinks] that* vs. *and he does not know that*) because the B-implication is subject to the strong contextual felicity constraint, but the *puisque*-implication is not (see Tonhauser et al. 2013).

and the local obligatory effect, falls into Tonhauser et al.'s 2013 class A, just like the existence of alternatives triggered by the focus particle *too*.

3. Two sources of not-at-issueness

The not-at-issueness of the two implications associated with *puisque*-clauses comes, I hypothesize, from two different sources: a syntactic one and a lexical one.

3.1 Syntactic source (the *puisque*-implication)

The reason why the *puisque*-implication is not at-issue is structural: *puisque*-clauses modify high phrases that are not at-issue, and we can assume that modifiers of not-at-issue content are not at-issue themselves.

Specifically, *puisque*-clauses like (1) or (4) can be argued to modify Evidential Phrases (EvidP, see Cinque 1999; Speas & Tenny 2003; Speas 2004, a.o.) and those like (2) or (5) to modify Speech Act Phrases (SAP, see Cinque 1999; Speas & Tenny 2003; Speas 2004; Haegeman & Hill 2013, a.o.).

This is first motivated by their meaning. As seen above, the former type of *puisque*-clauses provides evidence for the matrix proposition, and elements occupying EvidP such as adverbs like *apparemment* 'apparemment' or evidential markers specify the type of evidence for the statement expressed by their clause. Similarly, elements in SAP such as adverbs like *franchement* 'frankly' or speech act particles provide information about the speech act and the latter type of *puisque*-clauses expresses a reason for the speech act.

The second motivation for assuming that *puisque*-clauses sit in EvidP or SAP is their syntactic height: EvidP and SAP are high projections of the TP-space as indicated below (see Cinque 1999; Speas & Tenny 2003, a.o.), and *puisque*-clauses can be shown to scope as high.

(24) **SAP** > EvalP > **EvidP** > EpistP > T… > Asp… > V

Indeed, *puisque*-clauses cannot be outscoped by elements in Epistemic Phrases (EpistP) such as epistemic adverbs like *probablement* 'probably' or epistemic modals like *devoir* 'must' as illustrated in (25) and (26), respectively (cf. Rutherford 1970; Groupe λ-1 1975; Sæbø 1991, a.o.).[6] (27) however shows that *parce que*-clauses can be in the scope of these elements.

---

6. Conversely, (25a) and (26a) (under the acceptable reading) show that *puisque*-clauses can modify epistemic modals by specifying the content of indirect evidence (cf. von Fintel & Gillies 2010 on the relation between epistemic modals and indirect evidentiality).

(25) a. *#Lise est probablement partie puisque son manteau n'est pas là*
'#Liz probably left since her coat is not here'
[Intended: the absence of Liz's coat is a plausible evidence indicating that she left]

b. *#Lise est probablement partie puisque tu veux tout savoir*
'#Liz probably left since you must know everything'
[Intended: your willingness to know everything is a plausible reason for my asserting that Liz left]

(26) a. *#Lise doit être partie puisque son manteau n'est pas là*
'#Liz must have left since her coat is not here'
[Intended: the absence of Liz's coat is a plausible evidence indicating that she left]

b. *#Lise doit être partie puisque tu veux tout savoir*
'#Liz must have left since you must know everything'
[Intended: your willingness to know everything is a plausible reason for my asserting that Liz left]

(27) a. *Lise est probablement partie parce qu'elle était fatiguée*
'Liz probably left because she was tired'
[i.e. Liz's tiredness is a plausible reason why she left]

b. *Lise doit être partie parce qu'elle était fatiguée*
'Liz must have left because she was tired'
[i.e. Liz's tiredness is a plausible reason why she left]

The high scope of *puisque*-clauses is corroborated by the fact that unlike *parce que*-clauses as in (30) (cf. Rutherford 1970; Groupe λ-1 1975; Sæbø 1991; Johnston 1994, a.o.), they cannot be interpreted in the scope of the matrix negation,[7] and a matrix quantifier cannot bind into them, as exemplified in (28) and (29) respectively.

(28) a. *#Lise n'est pas partie puisque son manteau n'est pas là*
'#Liz did not leave since her coat is not here'
[Intended: the absence of Liz's coat is no evidence indicating that she left]

7. This has consequences for the projection tests presented in Section 2.1, given that implications are said to project only if their trigger is in the syntactic scope of the relevant operators: sentences like (28), where *puisque*-clauses are not in the syntactic scope of the negation, are not conclusive about the projection of their implication. But in the tests in Section 2.1, the *puisque*-clauses modify a different clause than the clause containing the operator (e.g. *ce n'est pas le cas que…puisque…* 'it is not the case that…since…'), and we have seen in Section 2.3 that evidential *puisque*-clauses can occur in embedded (attitude) clauses.

b. *#Lise n'est pas partie puisque tu veux tout savoir*
'#Liz did not leave since you must know everything'
[Intended: your willingness to know everything is not the reason for my asserting that Liz left]

(29) a. **[Aucune fille]$_i$ n'est partie puisque son$_i$ manteau n'est pas là*
'*[No girl]$_i$ left since her$_i$ coat is not here'

b. **[Aucune fille]$_i$ n'est partie puisque tu veux tout savoir sur elle$_i$*
'*[No girl]$_i$ left since you must know everything about her$_i$'

(30) a. *Lise n'est pas partie parce qu'elle était fatiguée*
'Liz did not leave because she was tired'
[i.e. Liz's tiredness is not the reason why she left]

b. *[Aucune fille]$_i$ n'est partie parce qu'elle$_i$ était fatiguée*
'[No girl]$_i$ left because she$_i$ was tired'

All these facts thus concur to show that *puisque*-clauses modify evidential or speech act phrases. Given that illocutionary force and evidential markers have been shown to be not-at-issue (Faller 2002; Murray 2010, a.o.), this implies that *puisque*-clauses modify not-at-issue content. Assuming that modifiers of not-at-issue content cannot be at-issue themselves, this explains why *puisque*-clauses, or more specifically *puisque*-implications, are not at-issue.

3.2 Lexical source (the B-implication)

The syntactic height of *puisque*-clauses thus accounts for the not-at-issueness of the *puisque*-implication, which is directly associated with *puisque*. The not-at-issueness of the smaller, B-implication, which corresponds to the content of the clause selected by *puisque*, is however lexically triggered: like factive verbs, *puisque* makes its complement clause factive. This directly explains why the B-implication belongs to the same class of projective content as complements of factive verbs (Tonhauser et al.'s 2013 class C, see Section 2.3).

The argument motivating this hypothesis comes from the behavior of *puisque*-clauses in attitude contexts. The main characteristic of factive verbs like *savoir* 'know' is that the speaker is committed to believe their complement (Karttunen 1974, among many others): in (31), not only Paul, but also the speaker believes that Liz left.

(31) *Paul sait que Lise est partie*
'Paul knows that Liz left'

The same holds with *puisque*-clauses. Recall from Section 2.3 that *puisque*-clauses are subject to the obligatory local effect, i.e. when *puisque*-clauses are in the syntactic scope of an attitude verb, the attitude holder must believe the content of their clause (the B-implication). Furthermore, *puisque*-clauses are subject to an

obligatory *global* effect:[8] when they are embedded in attitude contexts, the speaker too must believe the B-clause: in (32a), not only Paul, but also the speaker must believe that the neighbors' radio is on (the B-implication); this is corroborated by the fact that (32b) gives rise to a contradiction.

(32) a. *Paul pense que puisque leur radio est allumée, les voisins doivent être partis*
'Paul thinks that since their radio is on, the neighbors must have left'

b. *#Je ne pense pas que la radio des voisins soit allumée, mais Paul pense que puisque leur radio est allumée, ils doivent être partis*
'#I do not think that the neighbors' radio is on, but Paul thinks that since their radio is on, they must have left'

However, the speaker need not be committed to the evidential relation expressed by *puisque* (the *puisque*-implication): (33) is not contradictory.

(33) *Paul pense que puisque leur radio est allumée, les voisins doivent être partis. Moi, je ne pense pas qu'ils aient fait exprès de laisser la radio allumée en partant*
'Paul thinks that since their radio is on, the neighbors must have left. But I do not think that they left with the radio on on purpose'

The obligatory global effect is thus a specificity of the B-implication. This directly follows under the hypothesis that the subordinator *puisque* is factive, just like *savoir* 'know'.[9] Specifically, we can assume that *puisque* s-selects for arguments that denote facts. Assuming that *that*-clauses are either factive or propositional depending on their complementizer (Kratzer 2006, a.o.), this means that *puisque* selects factive *that*-clauses as represented in (34), while *parce que*-clauses select non-factive *that*-clauses: unlike the cause or explanation introduced by *because*, the evidence introduced by *puisque*-clauses is presented as a true fact.

(34) a. A *puisque* B ≈ A *puisque* the fact that B
b. [[*puisque* B]] = [[*puisque* $that_F$ B]][10]

---

8. This basically corresponds to what Amaral et al. (2007) call speaker-orientation. But unlike in the examples of conventional implicatures they provide, the speaker-orientation of the B-implication does not imply that it cannot be believed by the attitude holder too.

9. The standard view of factivity as a lexical trigger has been recently questioned by Beaver (2010) and Karttunen (2016), among others. But it remains unclear how double presupposition triggers like *puisque* could be treated under these pragmatic theories.

10. The factive complementizer $that_F$ remains silent with *puisque.* But it becomes overt (as *que*) in the case of coordination:

(iv) *Lise est partie, puisque son manteau n'est pas là et que je ne la vois pas*
'Liz left since her coat is not here and I cannot see her'

(35) $[\![that_F]\!]$ = λpλe. exemplifies (p)(e) or λp.ιe exemplifies (p)(e)
(*Kratzer 2006*)

4. Conclusion

In sum, the subordinator *puisque* can be added to the inventory of presuppositional triggers, which are associated with two types of not-at-issue implication: the relation it expresses (the *puisque*-implication) and the content of the clause it selects (the B-implication). These two not-at-issue implications have different properties: in particular, only the latter can be accommodated in contexts that do not imply it and exhibits factivity effects in attitude contexts. These differences follow, I have argued, from the fact that their not-at-issueness derives from two different sources: the (bigger) *puisque*-implication is not-at-issue because *puisque*-clauses modify high projections that are not at-issue; the (smaller, contained in the *puisque*-implication) B-implication is not-at-issue because *puisque*, which selects facts, lexically induces it.

Puisque thus exemplifies a case of double presuppositional trigger, which imposes different properties to its implications because of its syntactic position and its lexical entry. In the future, it would be interesting to test whether the constraints imposed by *puisque* on its two implications can be generalized to other double triggers: for instance, is it always the case that factivity characterizes the smaller implication (the one contained in the other)? do their different properties always come from different sources of not-at-issueness? An exploration of other subordinators in other languages[11] or in other domains (than causality) should shed light on this type of issues.

Funding

This work is supported in part by the NSF under grants 1424054 & 1424336.

Acknowledgments

Many thanks to the abstract reviewers and audience members of GR30, NELS47, LSRL47, SALT27, to the departments of Linguistics at Stony Brook, Rutgers, UMass Amherst, HUJ, NYU

11. See Scheffler 2008 for a study of German subordinators of causality (*denn* vs. *weil*) with respect to not-at-issueness.

and USC, and to the Harvard graduate students who took my seminar Ling205r in the Fall 2016, for helpful discussion and comments on work related to the topic of this paper.

References

Abrusan, Márta. 2016. "Presupposition Cancellation: Explaining the 'Soft–Hard' Trigger Distinction." *Natural Language Semantics* 24: 165–202. https://doi.org/10.1007/s11050-016-9122-7

Amaral, Patrícia, Craige Roberts, and E. Allyn Smith. 2007. "Review of *The Logic of Conventional Implicatures* by Chris Potts." *Linguistics and Philosophy* 30: 707–749. https://doi.org/10.1007/s10988-008-9025-2

Anand, Pranav, and Valentine Hacquard. 2013. "Epistemics and Attitudes." *Semantics and Pragmatics* 6 (8): 1–59. https://doi.org/10.3765/sp.6.8

Beaver, David. 2010. "Have you Noticed that Your Belly Button Lint Colour is Related to the Colour of Your Clothing?" In *Presuppositions and Discourse: Essays Offered to Hans Kamp*, ed. by R. Bäuerle et al., 65–99. Bingley: Emerald. https://doi.org/10.1163/9789004253162_004

Chierchia, Gennaro, and Sally McConnell-Ginet. 1990. *Meaning and Grammar: An Introduction to Semantics*. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. New York: Oxford University Press.

Ducrot, Oswald. 1983. "*Puisque*: essai de description polyphonique." *Revue Romane* 24: 166–185.

Faller, Martina. 2002. "Semantics and Pragmatics of Evidentials in Cuzco Quecha." PhD dissertation, Stanford University, Stanford.

Groupe λ-1. 1975. "Car, parce que, puisque." *Revue Romane* 10: 248–280.

Haegeman, Liliane, and Virginia Hill. 2013. "The Syntacticization of Discourse." In *Syntax and its Limits*, ed. by Raffaella Folli, Robert Truswell, and Christina Sevdali, 370–390. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199683239.003.0018

Johnston, Michael James Robert. 1994. "The Syntax and Semantics of Adverbial Adjuncts." PhD dissertation, University of California, Santa Cruz.

Karttunen, Lauri. 1974. "Presuppositions and Linguistic Context." *Theoretical Linguistics* 1: 181–194. https://doi.org/10.1515/thli.1974.1.1-3.181

Karttunen, Lauri. 2016. "Presupposition: What Went Wrong?." In *Semantics and Linguistic Theory (SALT)* 26, ed. by M. Moroney et al., 705–731. Ithaca, NY: Cornell University.

Kratzer, Angelika. 2006. "Decomposing Attitude Verbs." Talk given at the *Hebrew University of Jerusalem*.

Krifka, Manfred. 2014. "Embedding Illocutionary Acts." In *Recursion: Complexity in Cognition*, 59–87. Springer International Publishing. https://doi.org/10.1007/978-3-319-05086-7_4

Murray, Sarah. 2010. "Evidentiality and the Structure of Speech Acts". PhD dissertation, Rutgers University, New Brunswick.

Roberts, Craige. 1996. "Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics." *Working Papers in Linguistics 49: Papers in Semantics*. Ohio State University.

Rutherford, William. 1970. "Some Observations Concerning Subordinate Clauses in English." *Language* 46: 97–115. https://doi.org/10.2307/412410

Sæbø, Kjell Johan. 1991. "Causal and Purposive Clauses." In *Semantik – Semantics. Ein Internationales Handbuch zeitgenössischer Forschung – An International Handbook of Contemporary*

Research (HSK 6), ed. by Arnim von Stechow, and Dieter Wunderlich, 623–631. Berlin: de Gruyter.

Sæbø, Kjell Johan. 2016. "Information Structure and Presupposition." In *The Oxford Handbook of Information Structure*, Chapter 7.

Scheffler, Tatjana. 2008. "Semantic Operators in Different Dimensions." PhD dissertation, University of Pennsylvania, Philadelphia.

Simons, Mandy, Judith Tonhauser, David Beaver, and Craige Roberts. 2010. "What Projects and Why." In *Semantics and Linguistic Theory (SALT)* 20, ed. by Nan Li, and David Lutz, 309–327. Ithaca, NY: CLC.

Speas, Margaret. 2004. "Evidentiality, Logophoricity and the Syntactic Representation of Pragmatic Features." *Lingua* 114: 255–276. https://doi.org/10.1016/S0024-3841(03)00030-5

Speas, Margaret, and Carol Tenny. 2003. "Configurational Properties of Point of View Roles." In *Asymmetry in Grammar*, ed. by Anna Maria DiSciullo, 315–344. Amsterdam: John Benjamins. https://doi.org/10.1075/la.57.15spe

Tonhauser, Judith. 2012. "Diagnosing (not-)at-Issue Content." *Semantics of Under-represented Languages in the Americas (SULA)* 6: 239–254.

Tonhauser, Judith, David Beaver, Craige Roberts, and Mandy Simons. 2013. "Toward a Taxonomy of Projective Content." *Language* 89 (1): 66–109. https://doi.org/10.1353/lan.2013.0001

von Fintel, Kai. 2004. "Would You Believe it? The King of France is Back! Presuppositions and Truth Value Intuitions." In *Descriptions and Beyond*, ed. by Marga Reimer, and Anne Bezuidenhout, 315–341. Oxford: Oxford University Press.

von Fintel, Kai, and Anthony S. Gillies. 2010. "Must… Stay… Strong!." *Natural Language Semantics* 18 (4): 351–383. https://doi.org/10.1007/s11050-010-9058-2

Geminates and vowel laxing in Quebec French

Anaële Cuerrier & Charles Reiss
Concordia University

Laxing and harmony in Quebec French (QF) high vowels shows dialectal, register and perhaps even lexical variation. A recent proposal to handle some of the data (Poliquin 2006) contains a radical innovation to phonological theory concerning long-distance segment interactions. We question the necessity of such an account by pointing out that recognition of geminate sonorants in QF can explain some puzzling forms without recourse to new devices. Our account is supported by phonetic considerations, as well as by recognizing that the alternative both under and overgenerates lax vowels in surface forms.

Keywords: vowel harmony, French, Quebec French, phonology, laxing, syllable, geminate

1. Introduction

Chomsky (2007) poses the question "How little can be attributed to UG while still accounting for the variety of I-languages attained…?" as a way to characterize the 'bottom-up' approach to Universal Grammar. Studying such questions should converge with the 'top-down' approach, the goal of which is "to abstract from the welter of descriptive complexity certain general principles governing computation that would allow the rules of a particular language to be given in very simple forms, with restricted variety" (Chomsky 2000, p. 122).

Whenever a linguistic analysis contains a new formal device or relation, we need to ask ourselves, as bottom-up-ists, whether we really need such an innovation; and as top-down-ists, whether there are any other existing devices that allow us already to handle the phenomena under consideration. In this paper, we argue that the non-local, non-iterative vowel harmony analysis (NLNIVH) offered for Quebec French[1] by Poliquin (2006) is not a desirable enhancement of

1. Poliquin refers to Canadian French, but the subset of Canadian dialects he considers all are represented in the relevant properties in Quebec, where all the dialects we consider are spoken. So we refer henceforth to Quebec French (QF).

https://doi.org/10.1075/rllt.16.05cue

phonological theory from either top-down or bottom-up points of view. A simpler analysis arises from an appreciation of a basic property of QF ignored in many discussions: the language has geminate consonants.

2. Basics of Quebec French +High vowels

Quebec French has six high vowels on the surface:

(1) Surface high vowels of Quebec French[2]

$$i=\left\{\begin{array}{l}+\text{High}\\-\text{Back}\\-\text{Round}\\+\text{ATR}\end{array}\right\}\quad u=\left\{\begin{array}{l}+\text{High}\\+\text{Back}\\+\text{Round}\\+\text{ATR}\end{array}\right\}\quad y=\left\{\begin{array}{l}+\text{High}\\-\text{Back}\\+\text{Round}\\+\text{ATR}\end{array}\right\}$$

$$I=\left\{\begin{array}{l}+\text{High}\\-\text{Back}\\-\text{Round}\\+\text{ATR}\end{array}\right\}\quad U=\left\{\begin{array}{l}+\text{High}\\+\text{Back}\\+\text{Round}\\-\text{ATR}\end{array}\right\}\quad Y=\left\{\begin{array}{l}+\text{High}\\-\text{Back}\\+\text{Round}\\-\text{ATR}\end{array}\right\}$$

As is common in the literature, we treat +Atr as equivalent to Tense and – Atr as equivalent to Lax. A standard assumption is that the lax high vowels are all derived from tense ones. This appears to be true historically. Standard, international French (SF) has no lax high vowels on the surface. As we will see, it is not completely obvious that QF can be analyzed synchronically with just one series of high vowels.

As a first approximation, assuming that at least some lax high vowels are derived, there appears to be a rule of allophonic laxing in closed syllables:

2. We use the curly brackets of standard set theory to represent segments, for reasons given in Bale et al. (2014); Bale and Reiss (2018).

(2) Allophonic laxing in closed syllables

Open syllable	Closed syllable
petit [pt^{s}i] 'small-m.'	*petite* [pt^{s}ɪt] 'small-f.'
tout [tu] 'all-m.'	*toute* [tʊt] 'all-f.'
fumer [fyme] 'to smoke'	*(il) fume* [fʏm] 'he smokes'
bru [bry] 'daughter-in-law'	*brute* [brʏt] 'bully'

Morpheme alternants like [fym ~ fʏm] suggest a synchronic process that laxes high vowels in closed syllables. The pattern is also compatible with the general distribution in unrelated words like [bry, brʏt].

One complication in distribution is due to recent borrowings from English, which in the speech of all our informants, even those who would be considered monolingual francophones, lead to minimal pairs distinguished by tense *vs.* lax high vowels:

(3) Do recent borrowings show lexical lax high vowels?
- *team, pool* [tim, pul] *vs. ultime, poule* [ʏlt^{s}ɪm, pʊl] 'last, hen'

The borrowed *team* and *pool* do not have lax vowels, even though the words have high vowels in closed syllables. So, it may be necessary to recognize more than a single series of high vowels underlyingly. The non-alternating ones may be +ATR and the alternating ones could potentially be analyzed as unspecified for ATR. The ultimately correct analysis of these borrowings and the implications for native words is not explored by Poliquin, nor does it bear on our critique of his proposals.

A further complication that can be put aside for our modest purposes is the exact outcome of high vowels in syllables closed by a voiced fricative:

(4) High vowels before coda voiced fricatives:
- QF High vowels are long before voiced fricatives – there seems to be variation and uncertainty with respect to TENSE/LAX in this position
- *rouge, brise, écluse, dire, pur*
- [rʊːʒ, brɪːz, eklʏːz, d^{z}ɪːʁ, pʏːʁ] or [ruːʒ, briːz, eklyːz, d^{z}iːʁ, pyːʁ]
- 'red, breeze, canal lock, to say, pure'

Poliquin (p. 12) reports that these long vowels are tense on the surface, and he derives them *via* a so-called 'Duke-of-York gambit' (Pullum 1976) in which the vowels start out tense, undergo laxing, then become tense again, before being lengthened. In the data we elicited, these lengthened vowels seem to be lax, impressionistically. There may be dialectal variation in the realization of these lengthened vowels, but the issue does not bear on our primary concerns.

The last issue that complicates the distribution of high vowels in QF is vowel harmony. For many, but not all, speakers a high vowel in an open syllable can harmonize with a (derived) lax vowel in a following syllable. Consider some forms of the verb *cuisiner* 'to cook':

(5) Vowel Harmony
- *cuisiner* [kᵸi.zi.ne] 'to cook' *vs. (il) cuisine* [kᵸɪ.zɪn] 'he cooks'
- A lax vowels shows up in an **open** syllable, harmonizing with the vowel in the following closed syllable.

In [kᵸi.zi.ne] all the syllables are open, and as expected, the two high vowels surface as tense *i*. In the inflected form [kᵸɪ.zɪn], the word final vowel is in a closed syllable and it surfaces as lax ɪ, as expected. However, for some speakers, the first vowel of this form is also lax in harmony with the final vowel. Some speakers do not display harmony, so their pronunciation of *cuisine* is [kᵸi.zɪn].

If we restrict ourselves to two syllable words containing two high vowels and a final closed syllable (closed by anything other than a voiced fricative), there are two possible patterns in QF – either the vowel of the initial syllable is tense as is normal in an open syllable (QF2a), or it harmonizes with that of the final (QF2b):

(6) Patterns in disyllabics with high vowels and final closed syllable
- SF with no laxing: *(C)i.CiC.* [kᵸi.zin]
- QF2a with laxing, no harmony: *(C)i.CiC.* [kᵸi.zɪn]
- QF2b with laxing and harmony: *Cɪ.CɪC.* [kᵸɪ.zɪn]

However, once we consider longer words, the question arises of what happens to strings of three or more high vowels. Intuitively, one expects the following patterns for three syllable words with two high vowels in open syllables followed by a high vowel in a final closed syllable:

(7) Patterns in trisyllabics with high vowels and final closed syllable
- SF with no laxing: *i-i-i*
- QF3a with laxing, no harmony: *i-i-ɪ*
- QF with laxing and harmony:
  - QF3b with 'iterative' harmony: *ɪ-ɪ-ɪ*
  - QF3c with 'local' harmony: *i-ɪ-ɪ*

In other words, dialects with laxing and harmony will either apply the harmony across the board, yielding a string of three lax vowels (case b), or else, harmony will apply just to the first vowel to the left of the vowel subject to closed syllable laxing (case c). We agree with Poliquin that both dialects are attested.

Our disagreement with Poliquin is about another logical possibility that he identifies which becomes a focal point of his complex analysis. Poliquin claims that there are dialects that have both laxing and harmony, but generate an alternative to QF patterns (b) and (c):

(8) Poliquin's third option for a dialect with laxing and harmony: QF3d with NLNIVH: *ɪ-i-ɪ*

This pattern involves what Poliquin calls 'non-local, non-iterative vowel harmony (NLNIVH)'. Of course, it is hard to find examples of trisyllables or longer words with strings of high vowels that could potentially manifest this pattern, and Poliquin pools the data from the eight potential forms he identifies. The most common form, and the form he uses as a label for this word-shape (trisyllabic with three high vowels, two in open syllables and a final closed syllable) is *illicite* 'illicit'.

Our research agrees with Poliquin's findings of four QF pronunciations of the vowels in this word, in addition to the SF pronunciation, which is also attestable in Quebec because of the prestige of that dialect:[3]

(9) Five pronunciations of the vowels of *illicite*
SF3 *i-i-i* (no laxing, no harmony)
QF3a *i-i-ɪ* (laxing, no harmony)
QF3b *ɪ-ɪ-ɪ* (laxing, across-the-board harmony)
QF3c *i-ɪ-ɪ* (laxing, local harmony)
QF3d *ɪ-i-ɪ* (putative NLNIVH)

So, there is no dispute about the range of variation in the vowels manifested in a word like *illicite*. There are speakers who produce forms with no laxing (and thus no harmony), as in SF. There are speakers with laxing but no harmony, as in QF3a. There are speakers with laxing and across-the-board 'iterative' harmony, as in QF3b. There are speakers with laxing and 'local' harmony, as in QF3c. Finally, there are speakers with the vowels seen in QF3d, and our dispute with Poliquin lies with the explanation for such forms. In brief, Poliquin explains them as manifesting laxing and harmony, but with a non-iterative harmony process that targets the leftmost high vowel in the word, which in the case of *illicite*, is the initial vowel. Our findings suggest that in fact the pronunciation *ɪ-i-ɪ* for the vowels in *illicite* instead reflects the same dialect as manifested in QF3a, showing laxing without harmony. The difference in the two outcomes is a matter of the input representation and resulting syllable structure. The lax initial vowel in QF3d is a result of a

3. Some of our (non-linguist) informants made comments that displayed a high level of meta-linguistic awareness. For example, one speaker's comments suggested a perceived conflict between the formality of the word *illicite* and the non-prestige of laxing. He had trouble deciding whether the first vowel of *illicite* could be *ɪ*, but he was completely certain that he pronounced the non-formal word *illégal* with initial lax *i*. As we will see below, the NLNIVH model wrongly predicts a pronunciation [ilegal] with an initial tense *ɪ*, since the vowel appears to be in an open syllable with no high vowel with which to harmonize.

closed syllable created by a geminate /l:/. Such forms do not provide evidence of Poliquin's radical claim of NLNIVH.[4]

3. Three strikes against NLNIVH in *illicite*

We must reiterate that the data on vowel distribution in all of Canadian or Quebec French is quite complex, however, we are not ready to accept the NLNIVH analysis for the set of forms we call QF3d. We offer a conceptual argument, a logical argument, and an empirical argument, that all combine to cast doubt on the plausibility of NLNIVH.

3.1 'Long-distance' is not 'non-local'

Our conceptual argument is that the notion of non-locality in NLNIVH is anomalous in the linguistics literature. There are many cases of long-distance phenomena in both the syntactic and phonological literature, but they all refer to dependencies or interactions that *may* happen at a distance, not that *must* happen at a distance. For example, in languages that are reported to have long-distance anaphors, like Icelandic, a long-distance anaphor can cause ambiguity with respect to the antecedent (see Thráinsson 2007, for discussion and references):

(10) Long-distance anaphors
   a. **English** Local:
      Sigga$_i$ says that Maria$_j$ loves herself$_{*i/j}$
   b. **Icelandic** Long-distance (ambiguous):
      Sigga$_i$ segir að **Maria**$_j$ elski sig$_{i/j}$
   c. **Unattested** Obligatorily Long distance:
      Sigga$_i$ says that Maria$_j$ loves herself$_{i/*j}$

English anaphors like *herself* in (a) can only have a local antecedent such as the clausemate *Maria*$_j$. The Icelandic anaphor *sig* can be bound by either a local antecedent in the same clause, *Maria*$_j$, or a long-distance antecedent in a higher clause, *Sigga*$_i$. However, case (c) does not exist. There are no obligatory-long-distance binding phenomena where the antecedent *must* be non-local. The same holds for long distance effects like sibilant assimiliation or dissimilation. These relations/processes

4. Since geminates do not appear in SF, a form like QF3a could result from a speaker attempting a prestige pronunciation and suppressing non-prestige geminate /l:/, but 'failing' to suppress laxing in the final closed syllable. Our intuition is that this is as likely an explanation of QF3a forms as the idea that in some QF dialects there is no geminate /l:/ (either in this word, or generally).

can apply over a span, from one end of a word to the other, but they also can apply at a closer distance (for a survey, see Rose & Walker 2004). Poliquin's proposal that a relation holds between the first and last vowels of *illicite*, but not between the medial vowel and the final vowel would be quite extraordinary in the context of other well known long-distance interactions.[5] Long-distance relations seem to *include* local relations. This conceptual argument is not a proof that NLNIVH does not exist, but it highlights just how radical the proposal is. Maxime Papillon (p.c.) points out that NLNIVH appears to be an example of what has been called *first-last assimilation* in the computational phonology literature (Lai 2015; Chandlee & Heinz 2018), a pattern that this literature claims to be unattested for principled reasons.

3.2 NLNIVH overgenerates and undergenerates

As Poliquin points out, and as we illustrated above, disyllabics cannot distinguish between a dialect with long-distance harmony and local harmony; and in fact, disyllabics will also fail to distinguish among those two options and NLNIVH. All three dialects would treat Poliquin's class of words with the shape of *Philippe* the same: in each dialect we would see the pattern of [fɪlɪp] with two lax vowels.

Crucially, one word of this shape that Poliquin does not list is morphologically related to *illicite*: *licite* 'licit'. Among our informants, there is nobody who both produces *illicite* with *ɪ-i-ɪ* as predicted by NLNIVH, and also produces *licit* with *ɪ-ɪ*, which would be consistent with NLNIVH as well. Instead, speakers who produce *illicite* with *ɪ-i-ɪ* consistently produce *licite* with the vowels *i-ɪ*. So, for the disyllabic, they seem to have closed syllable laxing, but no harmony. Under the NLNIVH analysis, the trisyllabics and disyllabics would have to be subject to different grammars. Poliquin's own logic predicts a different outcome for *licite* in his proposed NLNIVH dialects than what we have seen attested. In brief, the NLNIVH account *overgenerates* lax vowels, since it wrongly predicts a lax initial vowel in *licite* for the relevant dialects (with attested [lisɪt] and [ɪlisɪt]). Disyllabics provide crucial evidence for the correct analysis of the trisyllabics.

In addition to overgenerating a lax vowel in the NLNIVH version of *licite*, Poliquin's account also *undergenerates* lax vowels – it fails to generate lax vowels that actually do surface. Poliquin mentions *illumine* 'illuminates' [ɪ.ly.mɪn], as another example of NLNIVH. However, his model predicts that the infinitive of this verb *illuminer* should be [i.ly.mi.ne], since there are no closed syllables in this form. Poliquin does not cite any pronunciation for *illuminer*, but our informants

5. Of course, Poliquin's long-distance harmony can be *accidentally* local, for example, in a two-syllable word like *Philippe*. The point is that the domain is defined as *starting* at the beginning of the word, at a distance from the triggering segment.

have a lax initial vowel in both forms of the verb. So, NLNIVH undergenerates the laxing of the initial vowel of the infinitive form. The attested lax initial [ɪ] of *illégal* is also undergenerated by the NLNIVH account, since there are no following lax high vowels with which the initial vowel could harmonize. NLNIVH wrongly predicts [ilegal].

3.3 Geminates to the rescue

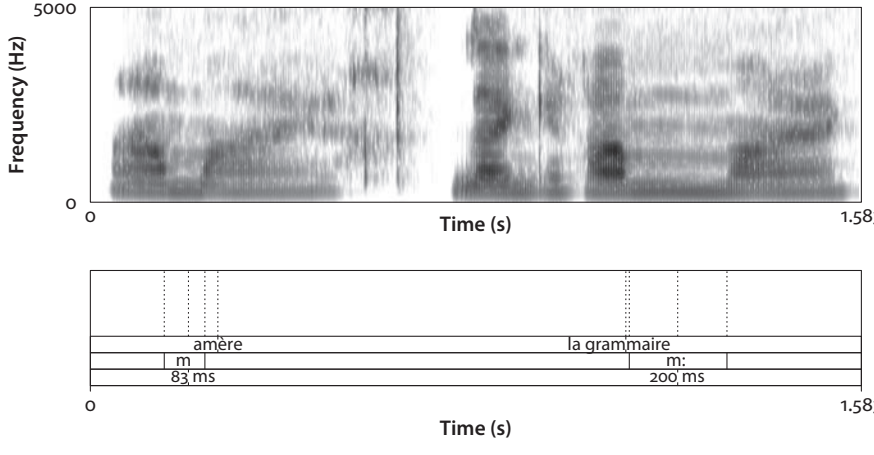
A simple, but underappreciated fact about QF phonology explains at least some of the unexpected lax vowels that Poliquin tries to account for with NLNIVH. This fact also explains the lax vowels that Poliquin fails to address, the ones NLNIVH undergenerates in *illuminer* and *illégal*. It turns out that many, if not all, dialects of QF have some geminate consonants, unlike Standard French (Morin 2010).[6] In QF geminates appear to be restricted to the sonorants /l,m,n/.

Tautomorphemic (underlying) geminates are present in words like *grammaire*, whereas a word like *amère* has a single [m]. Derived heteromorphemic geminates arise typically in clitic clusters, as in (11):

(11) Heteromorphemic gemnates in QF
- *elle apprend* [ælaprɑ̃] 'she is learning'
- *elle l'apprend* [æl:aprɑ̃] 'she is learning it'

Spelling is an unreliable indicator of gemination. The durational difference of single [m] and geminate [m:] in *amère* and *grammaire* is apparent in the spectrogram in (12):

(12) Spectrogram of *la grammaire, amère*



6. Walker (2001, Section 5.1) reports some geminates in SF, but they are irrelevant to our discussion.

The single [m] of *amère* is about 83 ms., whereas the geminate [m:] of *grammaire* is about 200 ms. Such an extreme difference, with the geminate more than twice as long as the single consonant, is not rare for sonorants (Local & Simpson 1999; Cohn et al. 1999).

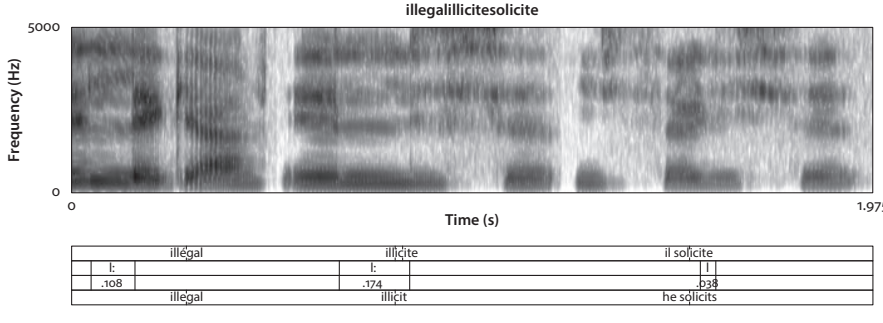
Poliquin does not recognize the existence of geminates in Canadian French, however, it is crucial to do so if we are concerned with phenomena that are sensitive to syllable structure. An obvious source of unexpected lax high vowels is the existence of syllables closed by the first half of a geminate.

Consider the spectrograms of three words, *illégal, illicite, il solicite* 'illegal, illicit, he solicits' in (13), all pronounced by the same speaker. The first word has a fairly long [l] of approximately 108 ms. The first vowel clearly sounds lax in this token, but since there are no other high vowels in the word, harmony cannot be relevant. As noted above, the NLNIVH account of laxing *undergenerates* in this case, because it does not predict the lax initial vowel of *illégal*. In fact, the vowel is lax because the following [l:] is a geminate, which closes the initial syllable – the correct analysis is [ɪl.le.gal]. This is just closed syllable laxing. The NLNIVH analysis cannot account for such facts.

The [l] of *illicite* is even longer, about 174 ms. The same explanation is available as for *illégal*. The lax initial vowel is due to laxing in a syllable closed by the first half of the long [l]. In this token of *illicite*, the medial vowel, in an open syllable, is tense, so this speaker appears to have no harmony from the final vowel. NLNIVH is not needed to account for the initial lax vowel – it is lax because it is in a closed syllable.

In *(il) solicite* the final vowel is lax because the syllable is closed by *t*. The preceding vowel is tense, which is consistent with what we see in *illicite* – this speaker again appears to have no harmony, which is why there is a tense medial vowel in *illicite*. In both words, medial *i* is in an open syllable, preceding a singleton [s].[7]

(13) Spectrogram of *illégal, illicite, il solicite*



7. Note that in QF there is no [l] pronounced at all in the subject clitic *il*.

The QF3d pronunciation of *illicite* with the vowels *ɪ-i-ɪ* reflects a dialect with no harmony, but a closed initial syllable: [ɪl.li.sɪt]. There is no reason to posit NLNIVH.

4. Conclusion

This paper is not an account of the messy distribution of lax high vowels in all Canadian French dialects (as well as all forms that reflect attempts to mimic SF and other prestige variants). Rather it is an argument against a particular radical enhancement of phonological theory that would allow for non-local non-iterative vowel harmony. Of course, it is impossible to prove that such phenomena do not exist in any language and that phonological theory will never be forced to accept the means to generate them.[8] However, our conceptual argument forces us to accept that NLNIVH is truly different from other long-distance phenomena we are familiar with. Our logical argument points out an error of prediction in that the NLNIVH approach predicts that speakers who have the *ɪ-i-ɪ* pattern in trisyllabics like *illicite* should have the *ɪ-ɪ* pattern in disyllabics like *licite*, contrary to our observations – NLNIVH overgenerates. Also, NLNIVH predicts that speakers who have an initial lax vowel in *illumine*, should have a tense initial in *illuminer*, again contrary to our observations. Like our mention of *illégal* above, this shows that NLNIVH undergenerates, since it predicts tense vowels (the absence of laxing), where the data manifests lax vowels. Under our account, the correct transcriptions of *illumine* and *illuminer* are [ɪl.ly.mɪn] and [ɪl.ly.mi.ne], both of which have lax vowels in an initial closed syllable.

So, NLNIVH both under- and overgenerates lax vowels under Poliquin's own assumptions. Finally, our empirical instrumental evidence attests to the presence of segmental durations that are compatible with the presence of phonological geminates, which in turn provide a straightforward account of some otherwise unexpected lax high vowels in QF.

Occam's Razor prevents us from accepting Poliquin's enhancement to the *computational* capacities of current models of phonology, an enhancement that would allow NLNIVH. Another recent PhD thesis, Bosworth (2011), also takes on the problem of the distribution of QF lax high vowels. Bosworth proposes an innovation to the *representational* capacities of phonological Universal Grammar, an entity she calls the *hypomora*. Bosworth, like Poliquin, fails to recognize the geminates of QF. Again, we think that radical solutions should not be considered

8. But see Lai (2015) and Chandlee and Heinz (2018) for suggestions that such gaps are principled.

until we have exhausted the resources of current models and paid fuller attention to a salient detail of the language under analysis. Although full diachronic and synchronic accounts of all forms produced by QF speakers would surely be complex and messy, there is no doubt that geminates are implicated in the distribution of high lax vowels.

We close with a somewhat philosophical point. We think that some readers will find our phonetic documentation of geminates and our arguments about over and undergeneration by NLNIVH more convincing than the conceptual argument that NLNIVH is unlike other long-distance phenomena. In terms of this paper, at least, it was our (rationalist) prejudices about what is a possible phonological process that led us to seek out further empirical facts and reject NLNIVH – our biases suggested that Poliquin's 'observations' could not be valid. We believe that this rationalist approach to inquiry is necessary and unavoidable – there are no facts that speak for themselves (Hammarberg 1981). This simple example from an accessible, fairly well-studied language suggests that one must be especially careful about accepting 'facts' that lead to radical theoretical innovations.

Acknowledgment

We are grateful to Justin Royer, Maxime Papillon and Michaël Gagnon for valuable suggestions.

References

Bale, Alan, and Charles Reiss. 2018. *Phonology: A Formal Introduction*. MIT Press.

Bale, Alan, Maxime Papillon, and Charles Reiss. 2014. "Targeting Underspecified Segments: A Formal Analysis of Feature Changing and Feature Filling Rules." *Lingua* 148: 240–253. https://doi.org/10.1016/j.lingua.2014.05.015

Bosworth, Yulia. 2011. "Weight and Feet in Québécois." Doctoral Dissertation, University of Texas, Austin.

Chandlee, Jane, and Jeffrey Heinz. 2018. "Strict Locality and Phonological Maps." *Linguistic Inquiry* 49: 23–60. https://doi.org/10.1162/LING_a_00265

Chomsky, Noam. 2000. "Language as a Natural Object." In *New Horizons in the Study of Language and Mind*, 106–133. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511811937.007

Chomsky, Noam. 2007. "Approaching UG from Below." In *Interfaces + re-cursion = Language?: Chomsky's Minimalism and the View from Syntax-Semantics*, ed. by Uli Sauerland, and Hans-Martin Gärtner, 1–24. Berlin: Mouton de Gruyter.

Cohn, Abigail C., William H. Ham, and Robert J. Podesva. 1999. "The Phonetic Realization of Singleton-Geminate Contrasts in Three Languages of Indonesia". *Proceedings of ICPhS* 14: 587–590.

Hammarberg, Robert. 1981. "The Cooked and the Raw". *Journal of Information Science* 3: 261–267. https://doi.org/10.1177/016555158100300603

Lai, Regine. 2015. "Learnable vs. Unlearnable Harmony Patterns". *Linguistic Inquiry* 46: 425–451. https://doi.org/10.1162/LING_a_00188

Local, John, and Adrian P. Simpson. 1999. "Phonetic Implemenation of Geminates in Malayalam Nouns". *ICPhS* 595–599.

Morin, Annick. 2010. "Diachrony and Synchrony of/l/Gemination in Québec French." In *Proceedings of the 2010 annual conference of the Canadian Linguistic Association*, ed. by Melinda Heijl, Online Proceedings: Canadian Linguistic Association.

Poliquin, Gabriel Christophe. 2006. "Canadian French Vowel Harmony." Doctoral Dissertation, Harvard University, Cambridge.

Pullum, Geoffrey K. 1976. "The Duke of York Gambit". *Journal of Linguistics* 12: 83–102. https://doi.org/10.1017/S0022226700004813

Rose, Sharon, and Rachel Walker. 2004. "A Typology of Consonant Agreement as Correspondence". *Language* 80: 475–531. https://doi.org/10.1353/lan.2004.0144

Thráinsson, Höskuldur. 2007. *The Syntax of Icelandic*. Cambridge University Press. https://doi.org/10.1017/CBO9780511619441

Walker, Douglas C. 2001. *French Sound Structure*, Volume 1. University of Calgary Press. https://doi.org/10.2307/j.ctv6gqrmv

Number as an adjunct in Romance

Sonia Cyrino & M.Teresa Espinal
University of Campinas / Universitat Autònoma de Barcelona

In this paper, we argue that Spanish (and other Romance languages such as Catalan), contrary to what has been believed up to now, patterns with languages such as Brazilian Portuguese and French. We present several different arguments to support our proposal that in the Romance languages we investigate plural marking is specified and interpreted on functional categories, namely on Determiners. We propose that in these languages plural marking is a syntactic adjunct to D (i.e., a categorized *d* root) by default, and to a categorized *n* root in marked cases. Manifestations of (plural) Number on other constituents within the nominal domain are to be considered solely as the output of morphophonological agreement or concord.

Keywords: pluralizer, adjunct, determiner, nominalizer, Romance

1. Introduction

Traditional grammars of Spanish (RAE 2009: 127) consider that Number is interpreted on the Noun, and the morphological (plural, singular) marking on the Determiner is the result of an agreement process, which cannot go missing, as illustrated in (1).

(1) *los* *libros* / **los* *libro* / **el* *libros* [Spanish]
the.PL book.PL the.PL book.SG the.SG book.PL
'the books'

The reason for this position is related to the classical theory that Number makes a clear semantic contribution to the interpretation of Nouns (Bartsch 1973; Hausser 1974; Bennett 1975; Schwarzschild 1996; Sauerland 2003, a.o.).

However, there are other Romance languages, such as Brazilian Portuguese (BrP) and French, in which, according to the literature (Delfitto & Schroten 1991, Bouchard 2002, Dobrovie-Sorin 2012, Cyrino & Espinal 2015), Number is encoded on D. One indication that this is correct is the following paradigm, where

https://doi.org/10.1075/rllt.16.06cyr

crucially the morphological marking for plural cannot occur only on the Noun in BrP.[1] As it is well-known, in French the distinction is instantiated only at the PF representation of the article: [lə] vs. [le].

(2) *os livros / os livro / *o livros* [BrP]
the.PL book.PL the.PL book.SG the.SG book.PL
'the books'

(3) *le livre / les livres* [French]
the.SG book.SG the.PL book.PL
'the book, the books.'

In this paper, we argue that Spanish (and other Romance languages such as Catalan and some Lunigiana dialects),[2] contrary to what has been believed up to now, patterns with languages such as BrP and French. Thus, we will support the hypothesis that in Spanish Number merges as an adjunct on D (or *d*), and only in marked cases it merges with a nominalized root (i.e., *n*).[3] We will argue that manifestations of plural marking on other constituents within the nominal domain are the result of morphophonological agreement or concord. Thus, we will account for some interesting cases of variation within the Romance paradigm.

2. Number is encoded and interpreted on D

In this section, we present several arguments, some of them from the literature, and others from our own, in support of our hypothesis that Spanish has Number specification and interpretation on functional categories, namely on D.

2.1 Arguments from the literature

i. As shown by Bouchard (2002: 42) for French, certain compound nouns are a V+N unit. Number is specified on the determiner, in order to convey the

1. The variation illustrated in (2) from BrP has been documented in several papers that focus on the sociolinguistics of BrP (Scherre 1994; Scherre & Naro 1998; Lopes 2006, among many others). Notice that although occurrences as **o livros* are ungrammatical in adult grammar, they occur at a first stage of plural marking in the acquisition of BrP. See in this respect Lopes (2006: 259).

2. Lunigiana is a geolinguistic domain extending over the borders between Liguria, Emilia and Tuscany.

3. We will hereby do not distinguish between a morphosyntactic PL feature and a semantic LATTICE feature, as proposed by Heycock and Zamparelli (2005).

distinction between reference to an atom or to a plurality of atoms that have the property 'can-opener'. The same applies to Spanish, as (4b) illustrates.

(4) a. ***l'*** *ouvre-boîte* / ***les*** *ouvre-boîte*
b. ***el*** *abrelatas* /***los*** *abrelatas*
the.SG open.can.PL the.PL open.can.PL
'the can-opener (singular and plural)'

ii. Longobardi (1994: 620) points out the need of distinguishing between coordination of DPs and coordination of NPs in Italian. In the former case (5a), each DP refers to a different person, and the verb is in plural; in the latter (5b), reference is made to only one person that has two properties ('secretary' and 'collaborator') and the verb is in singular. Therefore, manifestation of number on the verb is dependent on the existence of a coordination of DPs, which is structurally higher than a coordination of NPs. The same is possible in Spanish, as illustrated in (6):

(5) a. ***La*** *mia segretaria* ***e*** ***la*** *tua collaboratrice stanno/*
the my secretary and the your collaborator are/
**sta uscendo.*
is going out

b. ***La*** *mia segretaria* ***e*** *tua collaboratrice sta/ *stanno uscendo.*
the my secretary and your collaborator is/ are going out

(6) a. ***El*** *propietario* ***y*** ***el*** *gerente de la empresa viven/*
the.SG owner.SG and the.SG manager.SG of the firm live.PL
**vive en Andorra.*
lives in Andorra
'The owner and the manager of the firm live in Andorra.'

b. ***El*** *propietario* ***y*** *gerente de la empresa vive/ *viven*
the.SG owner.SG and manager.SG of the firm live.SG live
en Andorra.
in Andorra
'The owner and manager of the firm lives in Andorra.'

iii. It is possible to conjoin determiners in French and indicate number uncertainty (Bouchard 2002: 43). The same is possible in Spanish.[4] In English, by contrast, Plural is specified on the Noun as the translation of (8) shows.

4. A reviewer has pointed out to us the possibility that this might be a case of gapping. However, if this were the case, the example in (i) should be grammatical, contrary to facts.

(7) *Vous prendrez* ***le ou les*** *garçons que vous trouverez.*
you take.FUT the or the.PL boys that you find

(8) *Trae* ***el o los*** *diccionarios que encuentres.*
bring the.SG or the.PL dictionaries that find
'Bring the dictionary or dictionaries that you find.'

iv. Nominal ellipsis also shows that Number is specified in the Determiner not only in French and Walloon (Bouchard 2002), but also in Spanish (Torrego 1987; Kornfeld & Saab 2004).

(9) *Passe-moi* ***la verte****.*
give.me the green
'Give me the green one.'

(10) *Juan visitó a su tío y Pedro visitará a* ***los*** *de él.*
Juan visited to his uncle and Pedro visit.FUT to the.PL of he
'Juan visited his uncle and Pedro will visit his (uncles).'

Besides the arguments from the literature listed in this section, there are other reasons that lead to our main hypothesis that Number in Romance is encoded and interpreted on D. These additional arguments are presented in Section 2.2.

2.2 Additional arguments

v. In regular relatives (modified DPs), the complement of the D is not an NP, but a CP. On the one hand, in free relatives the *wh*-phrase of the embedded CP has been proposed to move to [Spec, DP], considering that the head of DP is a silent D (Caponigro 2002). This movement followed by agreement has been postulated to explain the plurality of the *wh*-word. Consider the minimal pair in (11) from Spanish.[5]

(11) a. ***Quién*** *llegue antes…*
whoever.SG arrive.SG before
'Whoever arrives before…'

b. ***Quienes*** *lleguen antes…*
whoever.PL arrive.PL before
'Whoever arrive before…'

(i) **Trae el diccionario*$_i$ *o los* e_i *que encuentres.*
bring the dictionary or the.PL that find.

5. A reviewer pointed out to us that the Spanish form *quienes* may be the result of agreement with a plural empty noun. However, it should be noted that no nominal form whatsoever is possible in this context, and therefore we cast aside this possibility.

On the other hand, in semi-free relatives (12) (de Vries 2000), the D is not silent, there is no Noun to trigger plural agreement on the verb, and Number can only be specified and interpreted on D.

(12) ***Los*** ***que*** *lleguen* *antes…*
the.PL that arrive.PL before…
'The (ones) that arrive before…

vi. *En*-anaphora in Catalan brings further evidence to our claim. The clitic *en* corresponds to a pro-N clitic (Déchaine & Wiltschko 2002), and it does not encode ϕ-features, in contrast to 3rd person clitics. As such, the structure in (14), corresponding to the answer in (13b), shows that, even though the cardinal may encode a plurality of atoms, no morphosyntactic Number is encoded on the pro-N.

(13) **Q.** *Que porta anells d'or*?
that wears rings of gold
A. *En porta* {*tres, un*}.
CL wears three one
'Does (s)he wear golden rings? (S)he wears {three, one}.'

(14) $[_{NP}$ ***En***$_i$] *porta* $[_{CardP}$ {*tres, un*} $[_{NP}$ $e_i]]]$

vii. Number on the 1st person pronoun (15a), and on the article preceding non-Spanish words, locutions and last names (15b) further support our claim (RAE 2009: 128–9).[6]

(15) a. ***Nos*** *el Rey* (…) *ordenamos y mandamos …*
we the king order.PL and command.PL…
b. ***los*** *mea culpa* / ***los*** *alto el fuego* / ***los*** *Escobar*
the.PL mea culpa the.PL ceasefire the.PL Escobar

---

6. A similar phenomenon to the one shown in (15a) has been observed in Greek. Lekakou and Szendröi (2012: 114) claim that in close appositives where a pronoun is syntactically combined with a full nominal DP the pronominal part is the unique head, as shown by the fact that verbal agreement is only possible with the pronoun. Consider (i).

(i) ***Emis*** *i glosoloji* *piname/* **pinane.*
we.NOM the linguists.NOM are.hungry.1PL/ are.hungry.3PL
'We linguists are starving/hungry.'

In contrast to Greek, in the Spanish example in (15a) no low boundary tone has to be pronounced at the end of *nos*, which suggests that *el rey* is not an apposition.
For an analysis of first and second person pronouns (as opposed to third person pronouns) as Ds, see Ritter (1991) and Déchaine and Wiltschko (2002).

In the former example we assume a structure with two determiners, where only the highest one, specified for plurality, c-commands and constrains the form of the verb. In the latter, overt instantiation of plurality on the noun is extremely marked (e.g., ?*los mea culpas*, **los meas culpa*).[7]

viii. Evidence from Afro-Bolivian Spanish (Delicado-Cantero and Sessarego (2011: 43–44) shows that number can be specified on determiners of all types: "as a rule, the nominal and the adjectival stems remain bare, so that plural marking is non-redundant".

(16) a. ***Ejes*** *buen amigo mayó*.
this.PL good.ø friend.ø old.ø
'These old good friends.'

b. ***Muchos*** *hombre boliviano*.
many.PL man.ø Bolivian.ø
'Many Bolivian men.'

Independent data from other Romance languages further support the hypothesis of morphosyntactic Number on D.[8] Of special relevance are the data from some Lunigiana dialects (Manzini & Savoia 2005; Pomino 2012; Cavirani 2018), and some creole varieties (Haitian creole, Ritter 1992, Déprez 2006; Afro-Brazilian Portuguese creole, Ribeiro & Cyrino 2012).

---

7. The plural of last names is possible if members of many families are referred to.

(i) *Los Morenos abundan en España.*
the.PL Moreno.PL are common in Spain
'The people named Moreno are common in Spain.'

8. Beyond the Romance paradigm Basque is a language where "The need to mark number on nouns explicitly by means of the plural marker [–k] forces the definite article [–a] to be also present" (Etxeberria 2014: 19). Consider the data in (i).

(i) a. *Anek goxoki-**a** jan zituen.*
Ane.ERG candy.D.SG.ABS eat AUX
'Ane ate the candy.'

b. *Anek goxoki-**ak** jan zituen.*
Ane.ERG candy.D.PL.ABS eat aux
'Ane ate (the) candies.'

(ii) a. **ikasle-**k***
student.PL

b. *ikasle-**a-k***
student.D.PL

Note that (iib) is ambiguous between a definite and an indefinite reading (the students, students). Information structure serves to disambiguate the readings. (U. Etxeberria, p.c.).

(17) a. [λa 'dɔna] Colonnata, new generations
the.PL woman
'the women'

b. [k-j-a 'brava 'dɔna] Filattiera
DEM.PL.FEM good.FEM woman
'those good women'

(18) *liv* ***yo*** Haitian creole
book the. PL
'the books'

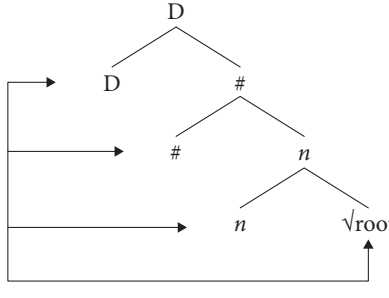
(19) ***Os*** *fio* Afro-Brazilian Portuguese creole
the.PL son
'the children'

Given all of this evidence, our proposal that Number is encoded and interpreted on D seems to be confirmed. Below, we present our analysis of plural marking in Romance.

3. Analysis

In order to analyze the data presented in Section 2 we start with Wiltschko's (2008: 688) proposal according to which plural comes in many guises and does not universally merge with nouns (see also Déprez 2005; Dobrovie-Sorin 2012; Mathieu 2014; Alexiadou 2016). In fact, as represented in (20), plural can either be merged on the root, on little n, on the functional # head, or on D, within the nominal domain.[9]

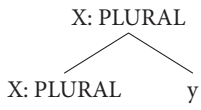
(20)



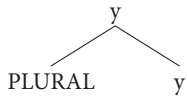
9. Note that Wiltschko's (2008) # is associated with one of two values: SINGULAR, spelled out as Ø in English and Romance, or PLURAL, spelled out by means of some allomorph. By contrast, Borer's (2005) # is associated with quantity, spelled out by means of cardinals or quantifiers.

Furthermore, according Wiltschko (2008: 688), plural can be merged either as a head (English) or as a modifier (Halkomelem). See (21).

(21) a.



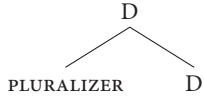
b.



The tree in (21a) is hypothesized for English, where the plural marker merges as a head and has the syntax of heads, whereas the tree in (21b) is hypothesized for Halkomelem, where a modifying plural marker shows the syntax of adjuncts.[10]

If we keep in mind the data presented in Section 2, we would like to put forward the hypothesis that in Romance the PLURALIZER is a modifier of D (or, alternatively, a modifier of a [Root + *d*]). Consider the tree in (22).[11]

(22)



This means that, in the default case, plural marking, i.e. the PLURALIZER, is a syntactic adjunct to D, spelled out on the D head and realized by Vocabulary Insertion as *-s* in both BrP and Spanish.

In (22), since the PLURALIZER is merged as an adjunct, it is syntactically opaque; hence, the newly formed object has the same label as its host (D).

Some consequences that follow from this analysis are the following:

- The PLURALIZER is morphosyntactically encoded and interpreted on D. Further instantiations of plural marking on nouns and adjectives within the nominal domain are the result of morphophonological agreement.
- The PLURALIZER is distinct from #:PL in that it does not introduce a feature to be valued at syntax.

---

10. This structural distinction is motivated by the fact that plural marking shows a set of properties and distribution that differ in English and Halkomelem. These properties are mainly: obligatory plural marking and obligatory agreement (only in English), and plural inside compounds and plural inside derivational morphology (only possible in Halkomelem). We will consider how Romance languages behave with respect to these properties of plural marking below.

11. See also Butler (2012) for the proposal of a DP-adjoined plural in Yucatec Maya.

This means that the operation of morphophonological agreement (i) is different from and is not the result of the syntactic operation Agree (as in Chomsky 2001), (ii) is optional, and (iii) takes place after Spell Out at PF. We hypothesize that some differences observed among Romance languages should be analyzed in terms of morphophonological agreement.

Consider first plural marking on nouns. We here repeat for convenience the BrP examples in (2), which contrast with the Spanish examples in (1).

(23) *os livros / os livro / *o livros* [BrP]
the.PL book.PL the.PL book.SG the.SG book.PL
'the books'

(24) *los libros / *los libro / *el libros* [Spanish]
the.PL book.PL the.PL book.SG the.SG book.PL
'the books'

It appears that plural marking is optional in BrP (French, Afro-Bolivian Spanish, some Lunigiana dialects and creole languages), but obligatory in Standard Spanish (Catalan and other Romance languages). We would like to claim that the PLURALIZER, by default, must modify D, and that morphophonological agreement is optional or obligatory, depending on the language. But, this property has nothing to do with syntax.

Consider next the phenomenon of agree within the nominal domain, as exemplified in (25) and (26).

(25) a. *as meninas bonitas* [BrP]
the.PL girl.PL pretty.PL

b. *as meninas bonita*
the.PL girl.PL pretty

c. *as menina bonita*
the.PL girl pretty
'the pretty girls'

(26) a. *las camisas blancas* [Spanish]
the.PL shirt.PL white.PL

b. **las camisas blanca*
the.PL shirt.PL white

c. **las camisa blanca*
the.PL shirt white
'the white shirts'

These examples show that Standard Spanish (in contrast to BrP and the Afro-Bolivian Spanish data in (16)) requires number agreement between the D and the

N within the nominal domain. This agreement requirement also applies to adjectives when they are present. However, as we have just said regarding the property of plural marking on nouns, we would like to address the phenomenon of agreement within the nominal domain by claiming that the PLURALIZER, by default, is a syntactic adjunct to D, and that morphophonological agreement within the nominal domain is obligatory or optional depending on the language. No syntactic principle or well-formedness condition controls this variation.

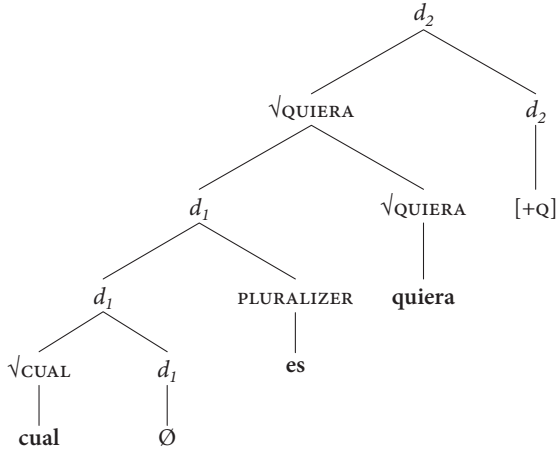
Let us next consider the possibility of having plural markers inside compounds. Regarding this property we should consider that both BrP and Spanish show complex determiners with plurality inside the compound, as exemplified in (27a) for BrP and (27b) for Spanish.

(27) a. *quaisquer*
which.PL.want
'whichever'

b. *cualesquiera*
which.PL.want
'whichever'

Following Harley's (2009) analysis of compounds, we assume that at the input structure for the Spanish complex D *cualesquiera* the root √CUAL is merged in complement position of a *d* functional head, and the *d*P it projects appears in object position of the root √QUIERA. Head movement of √CUAL into d_1, subsequently modified by the PLURALIZER, and later incorporated into the root √QUIERA derives the complex head [[[[√CUAL] d_1] PLURALIZER] √QUIERA], which later merges with a categorizing (quantifier-flavored) d_2, and head moves into it, creating the complex head [[[[[√CUAL] d_1] PLURALIZER] √QUIERA] d_2]. This complex head is finally realized by Vocabulary Insertion as *cualesquiera*. The output structure of these operations is represented in (28).

(28)



Also relevant to our analysis of Number as an adjunct is the consideration of data that include complex determiners, with articles and possessives. Consider in this respect the BrP paradigm in (29) and the Lunigiana data in (30) (from Cavirani 2018).

(29) a. (as) minhas coisas
the.PL my.PL thing.PL

b. (*a*) *minhas coisa*
the my.PL thing

c. (*as*) *minhas coisa*
the.PL my.PL thing

d. ??*as minha coisa*
the.PL my thing
'my things'

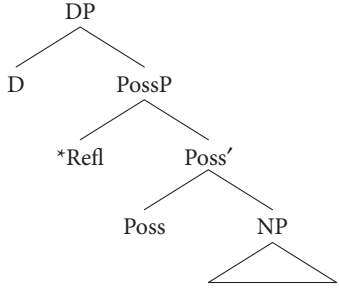
(30) a. Art.$_{PL.FEM}$ Poss.$_{PL.FEM}$ N.$_{FEM}$ Colonnata
[λa 'nɔʃtrja 'ka]
'our houses'

b. Art.$_{PL.FEM}$ Poss.$_{FEM}$ N$_{.FEM}$ Filattiera
[ja 'nɔstra so'rɛla]
'our sisters'

c. Art.$_{FEM}$ Poss.$_{PL.FEM}$ N.$_{FEM}$ Bagnone
[la 'nɔstrja 'skarpa]
'our shoes'

Following Despić (2015: 211), we consider that the possessor is in a construction preceded by a separate DP (see also Szabolcsi 1983; Kayne 1994). "This step is motivated by the fact that there are languages in which possessors are preceded by articles (e.g., [29] and [30]). The English prenominal possessor is also preceded by D, but this D is not pronounced in English". See the structure that Despić postulates in (31).

(31)



Our analysis of the data in (29) and (30) relies on the hypothesis that the real Determiner in this structure is the Possessive (like in English), and the article

may be omitted, as shown in BrP (29a,b,c). Since the Possessive is the head of the D-Poss complex unit, it may be modified by the PLURALIZER, as shown in (32) for (29b).

(32) $[_{DP}\ [_{D}$ a$]\ [_{PossP}\ [_{Poss}$ PLURALIZER $[_{Poss}$ minha $]]\ [_{NP}\ [_{N}$ coisa $]]]]$

Under this approach the optional plural marker on the definite article in (29a,c) and (30a,b) would be the result of morphophonological agreement.

Let us now go back to the figure in (20). Recall that Wiltschko (2008) defends the proposal that in English plural is merged as a head on #, whereas in Halkomelem plural is merged as a modifier on a √root. In this paper we have argued that in Romance plural marking is a modifier on D. The question that remains is whether there are any cases left that would support the need to consider that plural can also merge as a modifier of categorized nouns (i.e., *n*). In fact, such data exist. Consider the paradigms in (33), (34) and (35), which illustrate different sorts of nominal compounds.

(33) a. *coração* BrP
heart.SG
'heart'

b. *coraçãozinho*
heart.SG.diminutive
'little heart'

c. *corações*
heart.PL
'hearts'

d. *coraçõ**e**zinhos*[12]
heart.PL.diminutive.PL
'little heart'

(34) a. *tren**es*** *bala* Spanish
train.PL bullet
'bullet trains'

b. *gris**es*** *perla*
gray.PL pearl
'gray pearls'

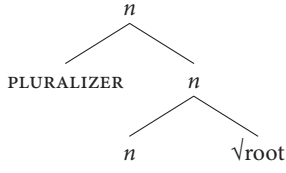
12. This example could be said to illustrate plurality inside derivational morphology. However, since it is widely accepted that inflection applies after derivation, we will consider that *coraçõezinhos* contains two pluralized nominals.

(35) a. *malsdecap* Catalan
ache.PL.of.head
'headaches'

b. *maldecaps*
ache.of.head.PL
'worries'

These data, as well as those examples where plural shows only on the noun in early stages of the acquisition of BrP (e.g., *o livros* 'the books', Lopes 2006), and in some Lunigiana dialects (e.g., Amegliese *a fantja* 'the girls', Cavirani 2018), suggest that in marked cases plural is a syntactic adjunct to a root categorized as *n*. This hypothesis, which we think accounts for those cases where plural marking is not made extensive to the determiner, is represented in (36).

(36)



To sum up, the existence of variation within a paradigm of languages that derive from Latin, a language that did not have articles, would be accounted for by postulating that plural marking is most commonly a syntactic adjunct to the D, and an adjunct to *n* only in marked, perhaps remnant, cases.[13]

4. Conclusion

In this paper we have argued that Number (PLURALIZER) in Romance is, by default, morphosyntactically encoded as an adjunct to the determiner root and morphophonologically instantiated as an option on other constituents within the DP. In marked cases Number (PLURALIZER) is morphosyntactically encoded as an adjunct to the categorized nominal root.

This conclusion supports the claim made in the literature that plural comes in many guises (Déprez 2005; Wiltschko 2008; Dobrovie-Sorin 2012; Mathieu 2014; Alexiadou 2016) and that it is not necessarily realized under a Division (#) head, as seems to be the case in English (cf. Borer 2005).

13. See Mathieu (2009) for an analysis of the emergence of D in the history of French.

Funding

Spanish MINECO FFI2014-52015-P FFI2017-82547-P; Generalitat de Catalunya to the Centre de Lingüística Teòrica 2017SGR634 CNPq, and ICREA Academia. (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil) 303742/2013-5.

Acknowledgments

We are very grateful to the audiences of several conferences (*IGG 2017* at Padova, *LSRL* 2017 at Delaware, *Encontro Nacional do Grupo de Trabalho de Teoria da Gramática 2017* at Salvador de Bahía, *Workshop on the Role of Parametric Variation at the Representation of Meaning 2017* at Bellaterra, *TEAM 2018* at Padova) for their comments and suggestions, as well as to two anonymous reviewers.

References

Alexiadou, Artemis. 2016. "The Syntax of Number." Talk given at the *Workshop on the Semantic Contribution of Det and Num: (In)definiteness, Genericity and Referentiality*. Universitat Autònoma de Barcelona, May 27–28.

Bartsch, Renate. 1973. "The Semantics and Syntax of Number and Numbers." In *Syntax and Semantics* 2, ed. by John Kimball, 51–93. New York: Seminar Press.

Bennett, Michael Ruisdael. 1975. *Some Extensions of a Montague Fragment of English*. Los Angeles, CA: University of California at Los Angeles dissertation, distributed by Indiana University Linguistics Club.

Borer, Hagit. 2005. *Structuring Sense. Vol I: In Name Only*. Oxford: Oxford University Press.

Bouchard, Denis. 2002. *Adjectives, Number and Interfaces. Why Languages Vary*. Amsterdam: Elsevier.

Butler, Lindsay K. 2012. "The DP-Adjoined Plural in Yucatec Maya and the Syntax of Plural Marking." Ms., University of Rochester, Rochester, NY.

Caponigro, Ivano. 2002. "Free Relatives as DPs with a Silent D and a CP Complement." In *Proceedings of WECOL 2000*, ed. by Vida Samiian, 140–150. Fresno: California State University, Department of Linguistics.

Cavirani, Edoardo. 2018. "The Distribution of Gender and Number in Lunigiana Nominal Expressions." In *Advances in Italian dialectology: Sketches of Italo-Romance Grammars*, ed. by Roberta D'Alessandro, and Diego Pescarini, 13–51. Leiden: Brill. https://doi.org/10.1163/9789004354395_003

Chomsky, Noam. 2001. "Derivation by Phase." In *Ken Hale: A Life in Language*, Michael Kenstowicz ed. by, 1–52. Cambridge, MA: MIT Press.

Cyrino, Sonia, and M. Teresa Espinal. 2015. "Bare Nominals in Brazilian Portuguese: More on the DP/NP Analysis." *Natural Language and Linguistics Theory* 33: 471–521. https://doi.org/10.1007/s11049-014-9264-6

Déchaine, Rose-Marie, and Martina Wiltschko. 2002. "Decomposing Pronouns." *Linguistic Inquiry* 33: 409–442. https://doi.org/10.1162/002438902760168554

Delfitto, Denis, and Jan Schroten. 1991. "Bare Plurals and the Number Affix in DP." *Probus* 3: 155–185. https://doi.org/10.1515/prbs.1991.3.2.155

Delicado-Cantero, Manuel, and Sandro Sessarego. 2011. "Variation and Syntax in Number Expression in Afro-Bolivian Spanish." In *Proceedings of the 13th Hispanic Linguistic Symposium*, ed. by Luis A. Ortiz-López, 42–53. Somerville, MA: Cascadilla.

Déprez, Viviane. 2005. "Morphological Number, Semantic Number and Bare Nouns." *Lingua* 115: 857–883. https://doi.org/10.1016/j.lingua.2004.01.006

Déprez, Viviane. 2006. "On the Conceptual Role of Number." In *New Perspectives on Romance Linguistics: Vol. I: Morphology, Syntax, Semantics, and Pragmatics*, ed. by Chiyo Nishida, and Jean-Pierre Y. Montreuil, 67–82. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.275.07dep

Despić, Miloje. 2015. "Phases, Reflexives and Definiteness." *Syntax* 18: 201–234. https://doi.org/10.1111/synt.12031

Dobrovie-Sorin, Carmen. 2012. "Number as a Feature." In *Functional Heads*, ed. by Laura Brugè, Anna Cardinaletti, Giuliana Giusti, Nicola Munaro, and Cecilia Poletto, 304–324. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199746736.003.0024

Etxeberria, Urtzi. 2014. "Basque Nominals: From a System with Bare Nouns to a System Without." HAL Id: artxibo-01071708 https://artxiker.ccsd.cnrs.fr/artxibo-01071708. https://doi.org/10.1075/la.219.14etx

Harley, Heidi. 2009. "Compounding in Distributed Morphology." In *The Oxford Handbook of Compounding*, ed. by Rochelle Lieber, and Pavol Stekauer, 129–144. Oxford, OUP Press. https://doi.org/10.1093/oxfordhb/9780199695720.013.0007

Hausser, Roland. 1974. *Quantification in Extended Montague Grammar*. Austin, TX: University of Texas dissertation.

Heycock, Caroline, and Roberto Zamparelli. 2005. "Friends and Colleagues: Plurality, Coordination, and the Structure of DP." *Natural Language Semantics* 13: 201–270. https://doi.org/10.1007/s11050-004-2442-z

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

Kornfeld, Laura, and Andrés Saab. 2004. "Nominal Ellipsis and Morphological Structure in Spanish." In *Romance Languages and Linguistic Theory 2002*, ed. by Reineke Bok-Bennema, Bart Hollebrandse, B. Kampers-Manhe, and Petra Sleeman, 183–198. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.256.11kor

Lekakou, Marika, and Kriszta Szendrői. 2012. "Polydefinites in Greek: Ellipsis, Close Apposition and Expletive Determiners." *Journal of Linguistics* 48 (1): 107–149. https://doi.org/10.1017/S0022226711000326

Longobardi, Giuseppe. 1994. "Reference and Proper Names." *Linguistic Inquiry* 25: 609–665. https://www.jstor.org/stable/4178880

Lopes, Ruth. 2006. "Bare Nouns and DP Number Agreement in the Acquisition of Brazilian Portuguese." In *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, ed. by Nuria Sagarra, and Almeida Jacqueline Toribio, 252–262. Somerville: Cascadilla Proceedings Project.

Manzini, Maria Rita, and Leonardo Maria Savoia. 2005. *I dialetti italiani e romanzi*. Alessandria: Edizioni Dell'Orso.

Mathieu, Eric. 2009. "From Local Blocking to Cyclic Agree." In *Determiners: Universals and Variations*, ed. by Jila Ghomeshi, Ileana Paul, and Martina Wiltschko, 123–158. Amsterdam: John Benjamins. https://doi.org/10.1075/la.147.04mat

Mathieu, Eric. 2014. "Many a Plural." In *Weak Referentiality*, ed. by Ana Aguilar-Guevara, Bert Le Bruyn, and Joost Zwarts, 157–181. Amsterdam: John Benjamins. https://doi.org/10.1075/la.219.07mat

Pomino, Natascha. 2012. Partial or complete lack of plural agreement. In *Inflection and Word Formation in Romance Languages*, ed. by Sascha Gaglia and Marc-Olivier Hinzelin, 201–229. Amsterdam: John Benjamins.

Real Academia Española (RAE). 2009. *Nueva Gramática de la Lengua Española*. Madrid: Espasa.

Ribeiro, Ilza, and Sonia Cyrino. 2012. "A estrutura do DP no crioulo de Cabo Verde e no PB de afro-descendentes." *Revista Letras UFPR* 84: 169–195.

Ritter, Elizabeth. 1991. "Two Functional Categories in Noun Phrases." *Syntax and Semantics* 25: 37–62.

Ritter, Elizabeth. 1992. "Cross-Linguistic Evidence for Number Phrase." *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 37: 197–218. https://doi.org/10.1017/S0008413100021952

Sauerland, Uli. 2003. "A New Semantics for Number." In *SALT XIII*, ed. by R. Young, and Y. Zhou, 258–275. Ithaca: Cornell University. https://doi.org/10.3765/salt.v13i0.2898

Schwarzschild, Roger. 1996. *Pluralities*. Vol. 61 (Studies in Linguistics and Philosophy). Dordrecht, Netherlands: Kluwer. https://doi.org/10.1007/978-94-017-2704-4

Scherre, Maria Marta Pereira. 1994. "Aspectos da concordância de número no português do Brasil." *Revista Internacional de Língua Portuguesa* 12: 37–49.

Scherre, Maria Marta Pereira, and Anthony Julius Naro. 1998. "Restrições sintáticas e semânticas no controle da concordância verbal em português." *Fórum Linguístico* 1: 45–71.

Szabolcsi, Anna. 1983. "The Possessor that Ran Away from Home." *The Linguistic Review* 3: 89–102. https://doi.org/10.1515/tlir.1983.3.1.89

Torrego, Esther. 1987. "On Empty Categories in Nominals". Ms., UMass. Boston.

de Vries, Mark. 2000. "Appositive Relative Clauses." *Linguistics in the Netherlands* 17: 221–231. https://doi.org/10.1075/avt.17.21vri

Wiltschko, Martina. 2008. "The Syntax of Non-Inflectional Plural Marking." *Natural Language and Linguistic Theory* 26: 639–694. https://doi.org/10.1007/s11049-008-9046-0

A sociophonetic investigation of Mexico City Spanish vowel reduction

Meghan Frances Dabkowski
University of Delaware, Department of Languages, Literatures & Cultures

This study is the first to use acoustic data to investigate linguistic and social factors conditioning phonetic vowel reduction in Mexico City Spanish. The acoustic analysis reveals that two complementary strategies are used for reduction: voice weakening and shortening. Voice weakening affects all vowels at relatively similar rates, and is favored by preceding voiceless consonants, following voiceless consonants, and following pauses, and is most frequent in post-tonic position. Shortening affects high vowels and /o/, and is favored by preceding and following voiceless consonants, and is most frequent in pre-tonic position and unstressed monosyllabic words. These results support an articulatory gestural overlap analysis, in which characteristics of articulatory timing and movement account for the variation.

Keywords: phonetics, phonology, sociolinguistic variation, Spanish dialectology, Mexican Spanish

Introduction

Spanish vowels have been traditionally described as remarkably stable, less subject to changes in quality, duration, and voicing than consonants. However, there is a growing number of studies showing that they do present more dialectal differences than initially assumed, varying in quality (e.g., Boyd-Bowman 1960; Holmquist 1998; Hualde 1989; Lope Blanch 1979; Oliver Rajan 2007), duration (e.g., Morrison & Escudero 2007; García 2016) and voicing (e.g., Delforge 2008a, 2008b, 2012; Lipski 1990; Lope Blanch 1963; Serrano 2006). Indeed, a notable exception to Spanish vowel stability is the vowel reduction observed in central Mexico, especially within the greater metropolitan area of Mexico City. The term "vowel reduction" is used as a blanket term here to denote various types of phonetic reduction, including shorter duration, devoicing or weakened voicing, and apparent deletion. While some attention has been given to the topic, there has been no systematic instrumental approach to vowel reduction in this variety until now. The current

https://doi.org/10.1075/rllt.16.07dab

study comprises the first systematic description and analysis of the phenomenon in Mexico City Spanish (henceforth MCS) that relies on acoustic data, thus contributing to a more thorough understanding of this process and informing our knowledge of reduction processes overall. The results lend further support to phonological models that consider gestures as fundamental.

While not especially common, unstressed vowel reduction, devoicing, and deletion processes have been documented in several world languages. In most languages studied, high vowels are the targets of devoicing and perceived deletion. As Gordon (1998) points out, high vowels are thought to be affected for two reasons: their inherently shorter duration, and the high tongue position need to produce them, which results in increased air pressure in the oral cavity and inhibits transglottal airflow, which is needed to maintain voicing. Voiceless consonantal environments are found to be the most conducive to reduction, as is domain-final position (e.g. Dauer 1980). With these findings in mind, we turn to the research on vowel reduction in Spanish.

Unstressed vowel reduction and devoicing has been documented and described more extensively for Andean Spanish (e.g. Delforge 2008a, 2008b, 2012) than for any other variety. Delforge (2008a) carried out the first comprehensive acoustic analysis of the phenomenon in any variety of Spanish. She divided reduced vowels into four categories, based on acoustic characteristics (Table 1), and concluded that in Cusco, unstressed vowel reduction primarily takes the form of devoicing. She found that /e/ and the high vowels are the most frequent targets, although targets vary according to word position, with word-medial high vowels and /e/ being most frequently affected, but no vowels significantly more likely to be affected word-finally. As for the surrounding segmental environment, Delforge concluded that devoicing cannot be entirely attributed to /s/ as previous studies suggest, but that voiceless consonants more generally contribute to the process. Regarding social factors, Delforge found that older speakers devoice most, and women do so least, both of which point to this feature as a change in progress in the Cusco region. She also found an interaction between age and socioeconomic status (SES) – in the youngest group, speakers from the lowest SES devoiced at rates double that of the highest SES.

To summarize findings on Andean Spanish, there are discrepancies regarding which vowels are the most likely to devoice, depending on the position in the word, but /e/ is consistently mentioned as a frequent target across varieties and contexts. Adjacent voiceless consonants, especially /s/, are also considered to be a major contributing factor to the variation observed, as well as post-tonic position, and phrase-final position.

With respect to this phenomenon in Mexico, the literature is surprisingly sparse, especially because, as Perissinotto notes (1975: 26), the weakening or loss

Table 1. Delforge's categories of reduced vowels (adapted from Delforge 2008a)

Category of reduction	Acoustic characteristics of vowel
Partially devoiced / shortened	Voice bar of 30 ms or less
Weakly voiced	Slightly longer but faint voice bar and lacking clear formant structure
Completely devoiced	No glottal tone present but some energy observed in first and second formants, syllable not temporally reduced
Apparently elided	Absence of formants and voice bar, syllable does appear to be shortened.

of unstressed vowels is one of the most salient characteristics of MCS pronunciation. Early descriptions of vowel reduction in Central Mexico (Boyd-Bowman 1952; Matluck 1952; Canellada & Zamora 1960) were mostly based on auditory impressions of spontaneous speech. All three studies agree that reduction takes place in all positions: word-initial, -internal and -final, pre- and post-tonic. Boyd-Bowman and Matluck, however, both mention post-tonic, and especially word-final position as the most frequent contexts for reduction. Boyd-Bowman also claims that compared to other vowels, /a/ resists the process. Adjacency to /s/ is also mentioned as a contributing factor, as is rapid speech.

Lope Blanch (1963) conducted the first systematic investigation of the phenomenon and provided quantitative analysis of his observations. He made audio recordings from which he identified four degrees of reduction, dividing them into discrete categories: (1) intense relaxation of the vowel, (2) an intermediate stage of reduction in which a weakened vowel is perceptible, (3) an intermediate stage in which a "light vocalic element" is difficult (but possible) to perceive, and (4) apparent elision. Devoicing is possible in 1, 2, and 3. Fewer than 2% of his reduced tokens were tonic vowels. Primary targets for reduction were mid vowels /e/ and /o/, and nearly 90% were adjacent to /s/. As for social patterning, the author finds none.

Serrano (2006) carried out a small-scale study to assess the continued validity of Lope Blanch's results, finding that linguistic factors predicting reduction are high word frequency, post-tonic position, and preceding and following voiceless consonants, as well as a possible role for intonational phrase-final position. Like Lope Blanch (1963), Serrano reported that overall, reduction does not seem to be conditioned by age or social class, although he does report a possible role for gender, stating that male speakers provide the most extreme cases of reduction and elision. This patterning led Serrano to suggest that MCS unstressed vowel reduction is a case of stable variation rather than a change in progress.

With these previous findings in mind, namely the higher rates of reduction in voiceless environments, the mid-vowel targets, and the apparent lack of social

patterning, I turn to the present study, which seeks to corroborate previous results using acoustic data and systematic statistical analysis, with the following research questions in mind:

1. What are the acoustic correlates of vowel reduction in MCS?
2. How frequently are MCS vowels reduced?
3. What linguistic and social factors condition voice weakening and shortening in Mexican Spanish?

Methodology

The data for this study were obtained from audio recorded sociolinguistic interviews. Interviews were approximately 30 minutes long and consisted of informal conversation resulting from open-ended questions asked by the author about topics such as family and area of residence. Participants were recorded in their homes, workplaces, or local cafes using a MacBook Pro and a head-mounted Plantronics USB microphone. A total of 73 participants were recorded, but only 20 of those, comprising 10 men and 10 women, ages 21–63, are included in the sample for this study. Most were from middle to upper-middle class backgrounds.

The acoustic analysis was informed by descriptions of the various categories of vowel reduction observed by Delforge (2008a) for Cusco Spanish, and Lope Blanch (1963) for Mexican Spanish. Beginning at the 10-minute mark of each recording, the first 160 vowel tokens were acoustically analyzed for each participant using Praat (Boersma & Weenink 2016), in order to include at least 150 per participant in the final analysis, given that some would be excluded. In the infrequent cases where there was excessive ambient noise, vowels from another later section of the interview with less noise were analyzed instead. All vowels that were not adjacent to another vowel or glide were measured, for a total of 3176 tokens. 123 of those were removed from further analysis due to extremely long duration (> 200 ms), and 84 additional creaky voice vowels were removed, for a final total of 3053 tokens. The duration of each vowel was segmented on one tier of a TextGrid, by marking the beginning of formant structure visible in the spectrogram, especially in the second formant, and marking the offset prior to evidence of constriction associated with the following consonant. The duration of voicing during the vowel was segmented on a second tier, along with a qualitative label or voicing category indicating whether the voicing was reduced and if so, what type of voice weakening was observed. Table 2 shows a summary of voicing categories observed along with their acoustic descriptions. An initial examination of the data showed that instead of simple presence or absence of voicing, many tokens show weak

or breathy voicing that can be partial in some cases.[1] Figure 1 below shows how vowels were segmented.

Table 2. Categories of voicing emerging from acoustic analysis

Category	Acoustic description
Fully voiced	Strong voice bar throughout vowel duration, periodic waveform
Partially weakened	Faint voice bar and low intensity in waveform for some portion of vowel duration
Completely weakened	Faint voice bar and low intensity in waveform throughout vowel duration
Partially devoiced	No voice bar, aperiodic waveform for a portion of vowel duration. Energy around F2/F3 sometimes visible in spectrogram.
Completely devoiced	No voice bar, aperiodic waveform throughout vowel duration. Energy around F2/F3 sometimes visible in spectrogram.
Apparently deleted	No acoustic evidence of vowel

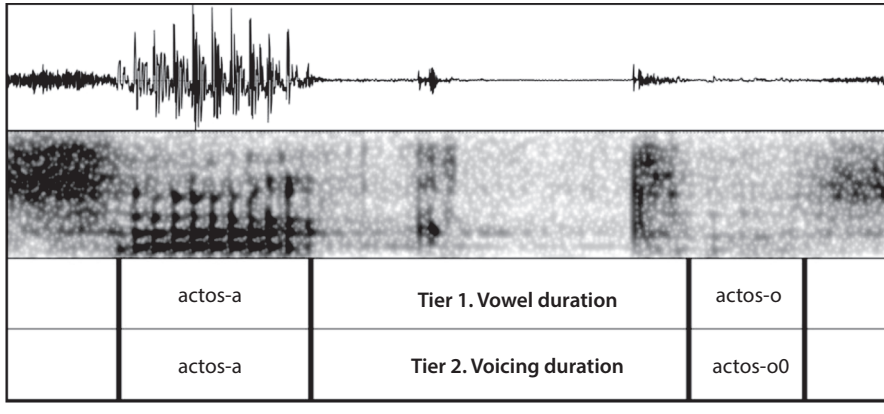


Figure 1. Praat TextGrid example of the measuring and labeling of vowels. The first vowel has full modal voicing and the second vowel has no voicing (hence the "0" label)

For the statistical analyses, the "partially weakened", "completely weakened", "partially devoiced", "completely devoiced", and "apparently deleted" categories were collapsed into one "weakened" category and compared to the "fully voiced"

1. Weak/breathy voicing is characterized by a lower intensity in the waveform, and a lighter voicing bar. Additionally, the presence of frication or energy in the higher frequencies is used as a way of distinguishing weakly voiced segments from devoiced segments when other aspects of the acoustic signal are not clear indicators of the presence or absence of voicing

category. These were the two levels for the dependent variable of *voice weakening*. However, this variable does not capture all types of reduction, as the acoustic analysis revealed many tokens that were fully voiced but extremely short, so a second model was developed in which *shortening* was the dependent variable. In order to quantify shortening as a categorical variable, i.e. to establish whether a token had been shortened or not, I calculated average vowel duration values in milliseconds for each speaker, by target vowel and stress[2]. Each token was then compared to the corresponding average, that is, each unstressed /a/ was compared to the average value for unstressed /a/ for that particular speaker, and tokens with a duration 50% or more shorter than the average[3] were coded as "shortened". Thus, the second dependent variable, *shortening*, had two levels: "shortened" and "not shortened".

Subsequently, I coded for independent variables found to predict vowel reduction in previous studies, namely target vowel, voicing of the preceding and following consonants, and position relative to lexical stress. I also included gender, to further evaluate Serrano's preliminary finding (2006) that men tend to produce more reduced vowels than women. Although vowel reduction was connected with speech rate in previous impressionistic studies of MCS, that analysis is beyond the scope of the current study but will be explored in future research. A summary of factors and their levels is shown in Table 3.

To explore the effects of these factors, I used the *glmer* function (Bates et al., 2015) in R (R Core Team, 2016) to develop two generalized linear mixed effects models using speaker as a random effect and the independent variables described above as fixed effects. The difference in the two models was their dependent variables – one model was run with *voice weakening* as the dependent variable, and a second with *shortening*. Based on prior research and my own observations during the acoustic analysis, several factors are hypothesized to influence vowel production in this variety. While it is unclear what effect, if any, social factors might have, my hypotheses regarding linguistic factors were the following: (1) vowels preceded by voiceless consonants are more likely to reduce; (2) vowels followed by voiceless

2. For example, speaker AS would have 10 averages computed, one for each of the 5 vowels (/a/, /e/, /i/, /o/, /u/), stressed and unstressed.

3. Setting the limit for determining whether a vowel was shortened at 50% shorter than average was a conservative way of capturing vowels that were significantly shorter than their corresponding averages, normalized by speaker. This is a more reliable measure than setting an arbitrary value as the limit: since faster speech rates result in shorter duration measurements, an arbitrary value could incorrectly identify vowels produced by a fast speaker as shortened when they are in fact in line with that speaker's average.

Table 3. Linguistic and social independent variables

Factor	Levels
target vowel	a, e, o, high (i, u)[4]
voicing of preceding consonant	voiced, voiceless, pause
voicing of following consonant	voiced, voiceless, pause
position relative to lexical stress	unstressed monosyllabic word, pre-tonic, post-tonic, stressed
speaker age	younger (21–35), older (38–63)
speaker gender	male, female

consonants or a pause are more likely to reduce; (3) mid vowels are more likely to reduce; and (4) vowels in post-tonic position are more likely to reduce.

Results

Overall rates of vowel reduction are low. The numbers reflected in Table 4 show the rates of voice weakening, shortening, reduction (composite of shortening and voice weakening), as well as the overlap between shortening and voice weakening, when considering the entire data set. Overlapped tokens are those that were both voice weakened and shortened, and as Table 4 shows, they make up less than 9% of all reduced tokens, and in fact only 1.2% of all tokens. From these percentages, we may conclude that voice weakening and shortening are two complementary strategies, each existing independently of the other.

Table 4. Overall reduction frequency rates

Weakening	Shortening	Weakening or shortening	Overlap
224/3053	248/3053	434/3053	38/434
7.34%	8.12%	14.22%	8.76%

Figures 2a–d below show voice weakening rates by target vowel, position relative to lexical stress, and voicing of the preceding and following consonants. Figure 2a

4. High vowels are the least frequent in Spanish, and due to their distribution across other variables in the data, they had to be collapsed together into one category.

shows relatively similar rates across target vowels. Figure 2b shows that voice weakening occurs at higher rates in vowels in post-tonic syllables than those in pre-tonic ones, unstressed monosyllabic words, and stressed syllables. Importantly, unstressed syllables account for much more of the reduction than stressed syllables, which is in line with cross-linguistic trends. Figure 2c shows a higher rate of voice weakening when the vowel is preceded by a voiceless consonant. Figure 2d shows a clear distinction between high rates of voice weakening in vowels followed by a pause, and quite low rates of reduction in vowels followed by a voiced consonant, with those followed by a voiceless consonant lying in the middle.

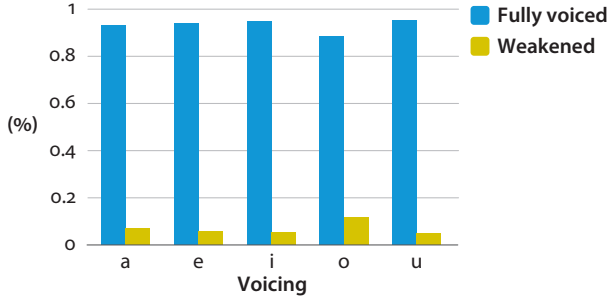


Figure 2a. Voice weakening rates by target vowel

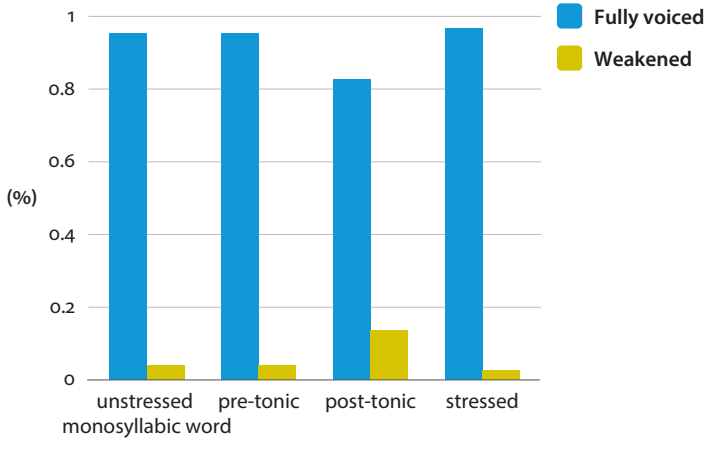


Figure 2b. Voice weakening rates by position relative to lexical stress

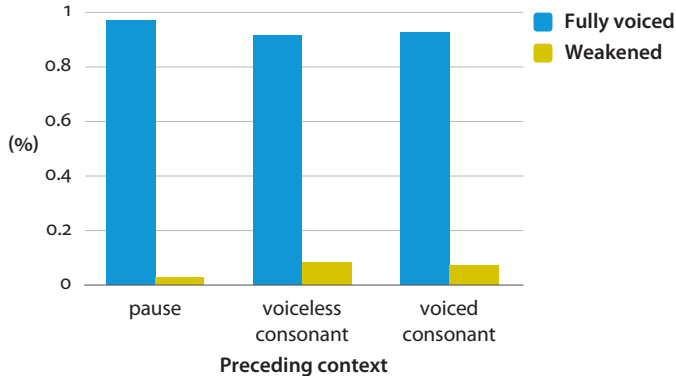


Figure 2c. Voice weakening rates by preceding context

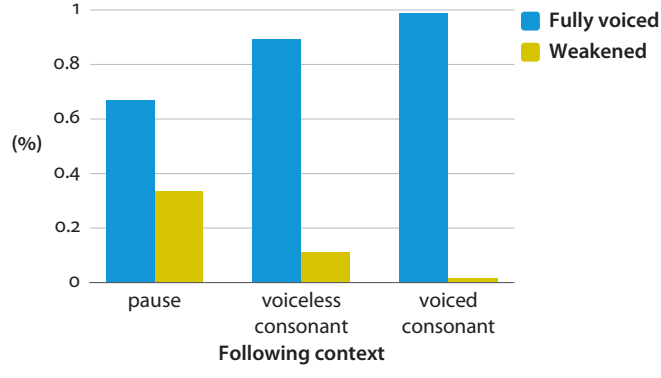


Figure 2d. Voice weakening rates by following context

Figures 3a–d below show shortening rates by target vowel, position relative to lexical stress, and voicing of the preceding and following consonants. Figure 3a shows the highest rates of shortening found in high vowels and /o/, as well as the lowest rate of shortening in /a/. Shortening rates are highest in unstressed monosyllabic words and pre-tonic syllables (Figure 3b). Preceding voiceless consonants result in the highest rates of shortening (Figure 3c), as do following voiceless consonants (Figure 3d), though a preceding pause seem to affect shortening differently than a following pause (Figure 3c, d).

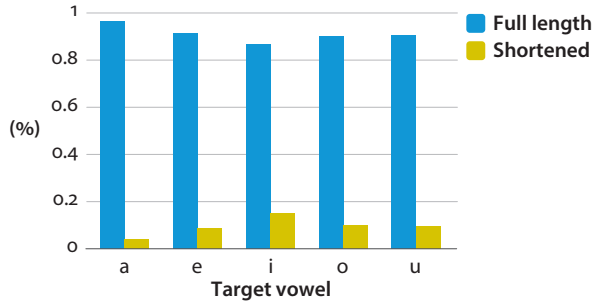


Figure 3a. Shortening rates by target vowel

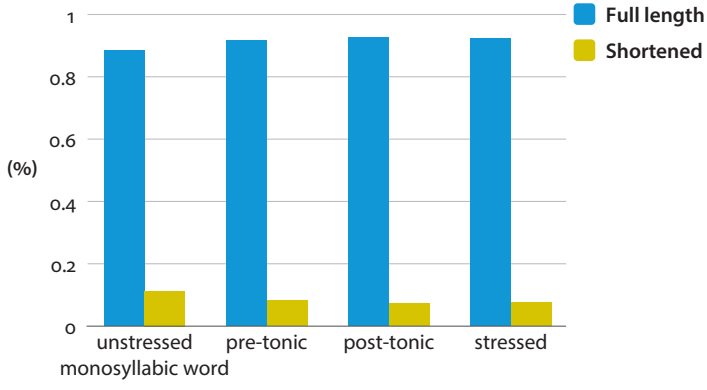


Figure 3b. Shortening rates by position relative to stress

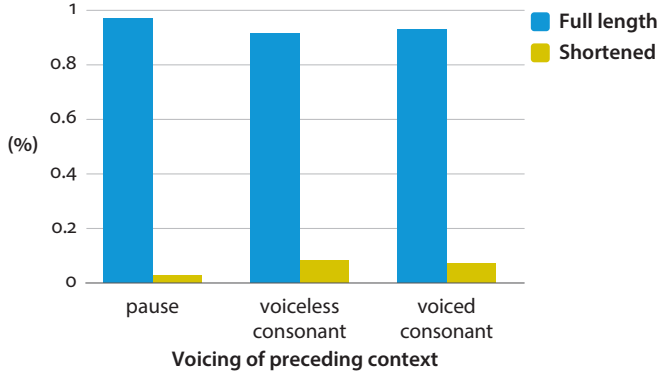


Figure 3c. Shortening rates by preceding context

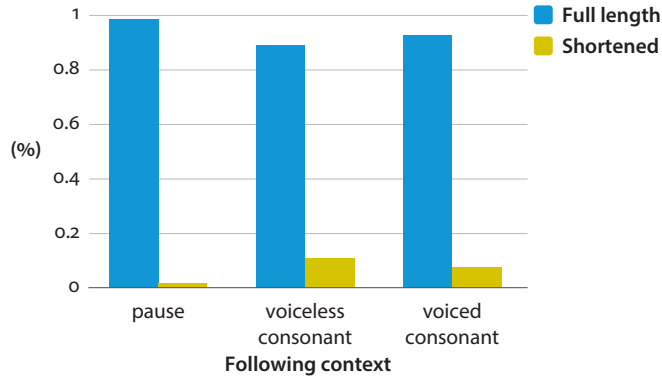


Figure 3d. Shortening rates by following context

Due to the low rates of voice weakening in stressed vowels (see Figure 2b) leading to several empty cells in the cross-tabulation matrix, the regression analysis of voice weakening was performed on a subset of the data including only unstressed vowel tokens (n = 1897). The results of the mixed-effects model using speaker as a random effect, and the *voice weakening* dependent variable, seen in Table 5, show that target vowel is not a significant predictor of voice weakening, but voicing of the adjacent segments, and the vowel's position relative to lexical stress are. Post-tonic position significantly favors weakening, as does a voiceless following consonant or pause and a voiceless preceding consonant.

Table 5. Factors contributing to voice weakening in subset of unstressed vowels

	Estimate	*SE*	*z*Value	*p*-value
(Intercept)	−2.85	0.19	−15.09	<0.001
Position relative to stress (reference level is average)				
Post-tonic	**0.45**	**0.09**	**4.80**	**<0.001**
Target vowel (reference level is average)				
/a/	−0.16	0.16	−1.05	0.29
/e/	0.14	0.15	0.90	0.37
/i, u/	−0.07	0.23	−0.30	0.77
/o/	0.10	0.14	0.67	0.50
Voicing of preceding segment (reference level is average)				
Voiceless consonant	**0.29**	**0.08**	**3.47**	**<0.001**
Voicing of following segment (reference level is average)				
Pause or voiceless consonant	**1.09**	**0.13**	**8.36**	**<0.001**

As above with voice weakening, because of the distribution of data across variables, the analysis of shortening on the subset of data included only unstressed vowels (n = 1897). The results of the mixed-effects model using speaker as a random effect, and the *shortening* dependent variable, indicate that target vowel, voicing of adjacent segments, and the vowel's position relative to lexical stress are all significant predictors of shortening. Table 6 shows that voiceless preceding segments and voiceless following consonants favor shortening. High vowels and /o/ are significantly more likely to be shortened, while /a/ is significantly less likely to do so. Vowels in monosyllabic unstressed words and pre-tonic position significantly favor shortening, while post-tonic position disfavors it.

Table 6. Factors contributing to shortening in subset of unstressed vowels

	Estimate	***SE***	***z*Value**	***p*-value**
(Intercept)	**−3.10**	**0.18**	**−17.22**	**<0.001**
Voicing of preceding segment (reference level is average)				
Voiced consonant/pause	**−0.82**	**0.10**	**−8.58**	**<0.001**
Position relative to stress (reference level is average)				
Monosyllabic unstressed word/				
pre-tonic syllable	**0.21**	**0.10**	**2.08**	**0.04**
Voicing of following segment (reference level is average)				
Voiceless consonant	**0.94**	**0.18**	**5.25**	**<0.001**
Target vowel (reference level is average)				
/a/	**−1.04**	**0.21**	**−5.04**	**<0.001**
/e/	0.15	0.14	1.06	0.29
/i, u/	**0.51**	**0.18**	**2.82**	**<0.01**
/o/	**0.38**	**0.15**	**2.59**	**<0.01**

Discussion

Most voice weakening (72%) takes place adjacent to one or more voiceless consonants, which suggests that voice weakening is in effect a type of coarticulation in which the vowel assimilates (or partially assimilates in the case of partially devoiced tokens or voice weakened tokens) to the preceding or following voiceless consonant. This assimilation can be explained in terms of the timing of glottal gestures (Browman & Goldstein 1990; Delforge 2008a, 2008b), specifically the overlapping of glottal opening gestures associated with the preceding consonant

and with the following consonant. That voiceless consonants also favor shortening may be further evidence for gestural overlap.

The finding that a following pause results in the most voice weakening can be understood by considering that the pause corresponds to intonational phrase-final, and often utterance-final position, which may be especially weak due to aerodynamic factors, that is, the fall in subglottal pressure throughout the time course of the utterance (Dauer 1980). The low rates of shortening before a pause are likely explained by phrase-final lengthening.

The result that rates of voice weakening were significantly higher in post-tonic syllables agrees with previous research. For Andean Spanish, Delforge (2008a) finds the most reduction in word-final position, which, due to the preponderance of paroxytonic words in Spanish, corresponds to post-tonic position. For Mexican Spanish, Serrano (2006) reports that the post-tonic syllable is the main locus for vowel reduction. I argue that these results can also be accounted for by gestural overlap, since post-tonic positions have been reported to be the weakest from an articulatory point of view, presenting reduced gestural magnitude and more overlap from adjacent gestures, resulting in more coarticulatory effects (e.g. Cole et al. 1999).

The results for the influence of the target vowel on weakening and shortening were partially in agreement with previous research. The literature suggests that while most languages target high vowels for reduction, Spanish targets mid vowels /e, o/ (Lope Blanch 1963; Delforge 2008a, b). However, in the current data, no vowel emerges as more likely than the others to favor voice weakening, indicating a possibility that in Mexican Spanish, voice weakening reduction is more generalized than in Andean Spanish, so that it is not vowel-dependent.

High vowels as well as /o/ all favor shortening. High vowels are already the shortest vowels intrinsically, so it may be that gestures associated with short vowels are prone to more extreme overlap with their surrounding consonants, and therefore subject to more shortening. Results also indicate that /a/ resists shortening, which is in line with expectations, as /a/ was found in previous studies to be the vowel least subject to reduction. The fact that target vowels are different for each type of reduction lends support to the idea that these are two distinct strategies for reducing vowels.

Age and gender were not chosen as significant by the best fit regression model. Given previous findings, and impressions that voice weakening and shortening are widespread, it is not surprising that neither a speaker's age nor their gender are predictors of voice weakening. It seems reasonable to conclude that the Mexico City vowel reduction is not a change in progress, but rather is a case of stable variation. This is quite different from the social situation in Cusco, where vowel devoicing is receding and current trends indicate that it is likely to disappear entirely (Delforge 2012).

Conclusions

The results of this study indicate that vowel reduction in Mexico City Spanish is characterized by two complementary strategies: voice weakening and extreme shortening. Voice weakening is conditioned by preceding voiceless consonants, following voiceless consonants, and following pauses, and is most frequent in post-tonic position. Shortening is conditioned by preceding voiceless consonants, following voiceless consonants, and is most frequent in pre-tonic position and unstressed monosyllabic words. Only a small number of tokens are both voice weakened and shortened, leading to the conclusion that they are distinct strategies that both contribute to vowel reduction in MCS.

Future research will determine the role of additional linguistic factors on both voice weakening and shortening, especially those of phrase position, word position, place and manner of surrounding segments, speech rate, and frequency. In addition, it will be possible to tease apart the different types of voice weakening as more data is analyzed. Finally, a possible role for socioeconomic status will be explored.

References

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, Paul, and David Weenink. 2016. *Praat: Doing Phonetics by Computer* [Computer program]. Version 6.0.14, retrieved 11 February 2016 from http://www.praat.org/.

Boyd-Bowman, Peter. 1952. "La pérdida de vocales átonas en la altiplanicie mexicana." *Nueva Revista de Filología Hispánica* 6 (2): 138–140. https://doi.org/10.24201/nrfh.v6i2.3163

Boyd-Bowman, Peter. 1960. *El habla de Guanajuato*. Santiago de Compostela: Imprenta universitaria.

Browman, Catherine, and Louis Goldstein. 1989. "Articulatory gestures as phonological units." *Phonology* 6 (2): 201–251.

Canellada de Zamora, María Josefa, and Alonso Zamora Vicente. 1960. "Vocales caducas en el español mexicano." *Nueva Revista de Filología Hispánica* 14: 222–241.

Cole, Jennifer, José Ignacio Hualde, and Khalil Iskarous. 1999. "Effects of Prosodic and Segmental Context on/g/-Lenition in Spanish." *International Linguistics and Phonetics Conference* 2: 575–589.

Dauer, Rebecca. 1980. "The Reduction of Unstressed High Vowels in Modern Greek." *Journal of the International Phonetic Association* 10: 17–27. https://doi.org/10.1017/S0025100300002073

Delforge, Ann Marie. 2008a. "Unstressed Vowel Reduction in Andean Spanish." *Laboratory Approaches to Spanish Phonology*, ed. by Laura Colantoni, and Jeffrey Steele, 107–124. Somerville, MA: Cascadilla Proceedings Project.

Delforge, Ann Marie. 2008b. "Gestural Alignment Constraints and Unstressed Vowel Devoicing in Andean Spanish." *West Coast Conference on Formal Linguistics*, ed. by Charles B. Chang, and Hannah J. Haynie, 147–155. Somerville, MA: Cascadilla Proceedings Project.

Delforge, Ann Marie. 2012. "'Nobody Wants to Sound Like a Provinciano': The Recession of Unstressed Vowel Devoicing in the Spanish of Cusco, Perú." *Journal of Sociolinguistics* 16 (3): 311–335. https://doi.org/10.1111/j.1467-9841.2012.00538.x

García, Miguel. 2016. "Sobre la duración vocálica y la entonación en el español amazónico peruano." *Lengua y Sociedad* 14 (2): 5–29.

Gordon, Matthew. 1998. "The Phonetics and Phonology of Non-Modal Vowels: A Cross-Linguistic Perspective." *Berkeley Linguistics Society* 24: 93–105. https://doi.org/10.3765/bls.v24i1.1246

Holmquist, Jonathan C. 1998. "High-Lands High Vowels: A Sample Of Men's Speech in Rural Puerto Rico." *Papers in Sociolinguistics: NWAVE-26 a l'Université Laval*, ed. by Claude Paradis, Diane Vincent, Denise Deshaies, and Marty LaForest, 73–79. Quebec: Université Laval.

Hualde, José Ignacio. 1989. "Autosegmental and Metrical Spreading in the Vowel-Harmony Systems of Northwestern Spain." *Linguistics* 27: 773–805. https://doi.org/10.1515/ling.1989.27.5.773

Lipski, John. 1990. "Aspects of Ecuadorian Vowel Reduction." *Hispanic Linguistics* 4 (1): 1–19.

Lope Blanch, Juan M. 1963. "En torno a las vocales caedizas del español mexicano." *Nueva Revista de Filología Hispánica* 17: 1–20.

Lope Blanch, Juan M. 1979. "Sobre el tratamiento de -e, -o finales en el español de México." *Investigaciones sobre dialectología mexicana*, ed. by Juan M. Lope Blanch, 35–40. México, D. F.: Universidad Nacional Autónoma de México.

Matluck, Joseph. 1952. "La pronunciación del español en el valle de México." *Nueva Revista de Filología Hispánica* 6 (2): 109–120. https://doi.org/10.24201/nrfh.v6i2.248

Morrison, Geoffrey, and Paola Escudero. 2007. A Cross-Dialect Comparison of Peninsula-and Peruvian-Spanish Vowels." *16th International Congress of Phonetic Sciences*, 1505–1508. Saarbrücken: University of Saarbrucken.

Oliver Rajan, Julia. 2007. "Mobility and its Effects on Vowel Raising in the Coffee Zone of Puerto Rico." *Third Workshop on Spanish Sociolinguistics*, ed. by Jonathan Holmquist, Augusto Lorenzino, and Lotfi Sayahi, 44–52. Somerville, MA: Cascadilla Proceedings Project.

Perissinotto, Giorgio Sabino Antoni. 1975. *Fonología del español hablado en la ciudad de México: Ensayo de un método sociolingüístico*. México D.F.: El Colegio de México.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. URL: https://www.R-project.org/.

Serrano, Julio. 2006. "En torno a las vocales caedizas del español mexicano: una aproximación sociolingüística." *Los líderes lingüísticos. Estudios de variación y cambio* 3: 7–59.

Computational quantitative syntax

The case of Universal 18

Kristina Gulordava & Paola Merlo
University of Geneva

Accounting for the constraints on the possible word orders of a sentence in a language and across the world languages is a core challenge for syntactic theory. In the spirit of computational quantitative syntax, in this paper we present quantitative evidence about Universal 18. We show that corpus data confirms a dispreference for the word order combination where adjectives precede but numerals follow the nouns (Adj-N and N-Num). We then investigate if this dispreference is better explained as a constraint expressed at the level of the dominant orders or at the level of individual structures. Corpus counts support the latter interpretation. Finally, we propose a formal model of how this bias against Adj-N-Num orders can be integrated in the grammar.

Keywords: Universal 18, quantitative syntax, corpus counts, adjective-noun order, noun-numeral order, treebanks, Latin, Ancient Greek, modelling

1. Introduction

Languages arrange constituents in a sentence in very different ways. Consider the same sentence in English, Irish and Persian, illustrated in (1)–(3).

(1) James saw her on top of the hill. English

(2) Chonaic Séamus ar bharr an chnoic í. Irish
saw Séamus on top of hill her
'Séamus saw her on top of the hill'

(3) Ali dokhtari ra dar balaye tapeh did. Persian (Farsi)
Ali girl *acc* on top hill saw
'Ali saw a girl on top of the hill'

Looking across all documented languages of the world, it is apparent that not all possible word order arrangements are attested. Moreover, it has long been known that there are correlations between the orders of certain grammatical categories. For example, if, in a language, the verb precedes its complement, as in 'saw a girl',

https://doi.org/10.1075/rllt.16.08gul

then very frequently the adposition will also precede the noun, as in 'on the hill' (Dryer 1992). Accounting for the constraints on the possible word orders of a sentence is a core challenge for syntactic theory.

One of the goals of computational quantitative syntax is to determine the status that frequency plays in this context. The current work investigates the connection between the quantitative aspects of word order variation across languages and the quantitative aspects of word order variation within a language. We take as our starting point the empirical generalisations on word order typology known as word order universals (Greenberg 1963). We study the corresponding patterns in language-internal variation by looking at syntactically-annotated corpora of various languages. Specifically, we investigate in what way Universal 18 – which restricts the combination of word orders between adjective and noun on the one hand, and numeral and noun on the other – is a constraint pertinent to the language system.

Universal 18 states a constraint on cross-linguistic distributions of the combination of two NP-internal orders: the order between adjective and noun and the order between numeral and noun.

> When the descriptive adjective precedes the noun, the demonstrative and the numeral, with overwhelmingly more than chance frequency, do likewise.
>
> (Greenberg 1963, p. 67–68)

As can be seen, this is already a statistical universal, although its quantification is not precisely numerical. We study the quantitative aspects of this statement, in the framework of current theoretical proposals. Specifically, we analyse two types of quantified observations about word order: *token-level* and *type-level* observations. Token-level observations correspond to the individual instances of word order realisation in a language. All instances of word order alternation at the token level constitute the language-internal variation. Type-level observations correspond to the word order properties of a language as observed in the grammar or in the typological descriptions. For example, the notion of "most frequent word order" (see below, footnote 4) is a type-level property of a language. Typological variation describes the distribution of languages by types, according to such properties. The formulation of Universal 18 above is ambiguous between these two interpretations. The statement reported above could mean that the order of adjective and noun and the order of numeral and noun are very frequently harmonious across the grammars of the world languages or it could mean that if we look at the actual sentences and phrases in a language the cooccurrence of prenominal adjective and postnominal numeral in the same phrase is dispreferred.

We exploit two different types of linguistic resources – syntactically annotated corpus data and typological data collected from large databases – to collect the

quantitative distributions of the word order variation expressed by Universal 18, within a language and across languages. In the next section, we discuss new typological data collected from large-scale typological resources. The rest of the paper will concentrate on corpus counts and language-internal variation.

2. Typological data on Universal 18

The World Atlas of Language Structures (WALS) (Dryer & Haspelmath 2011)[1] and the Syntactic Structures of the World's Languages (SSWL) databases[2] are large-scale typological resources that allow us to revisit Universal 18. Assuming that a language has categories such as numerals and adjectives, there are four possible combinations of word orders between adjective and noun and numeral and noun in a noun phrase: L1 – Adjective-Noun & Numeral-Noun; L2 – Adjective-Noun & Noun-Numeral; L3 – Noun-Adjective & Numeral-Noun; L4 – Noun-Adjective & Noun-Numeral.

According to Greenberg's generalisation, the languages of type L2 which exhibit both word orders Adj≺N & N≺Num are very infrequent. This tendency has been shown to be very robust and has been attested in a large sample of languages of the world (Hawkins 1983; Rijkhoff 1998; Hurford 2003). New supporting evidence can be found by consulting the databases. Table 1 presents the number of languages of each word order type as found in the World Atlas of Languages (WALS).[3] Out of four possible language types, the most infrequent one (4%) is Adj≺N & N≺Num type which confirms the generalisation of Universal 18.

Table 1. Counts of languages according to their adjective-noun and numeral-noun word order types, as presented in WALS

	Num≺N	**N≺Num**	**No dominant order**	**Total**
Adj≺N	251	37	12	300
N≺Adj	168	510	35	713
No dominant order	35	22	13	70
Total	**454**	**569**	**60**	**1083**

1. http://wals.info/

2. http://sswl.railsplayground.net/

3. These are counts of all languages in the database, without sampling by genera. See, for example, Hurford (2003) for the counts in an accurately constructed representative sample of languages of the world.

Generative syntacticians have typically paid more attention to the different possible orders within a language, rather than to the dominant order. Using the Syntactic Structures of the World's Languages (SSWL), we can show that a constraint against Adj≺N & N≺Num – Universal 18 – is also observed for combinations of word order variation types, that is, the different grammatically possible word orders that a given language allows. The SSWL database lists grammatical and ungrammatical morpho-syntactic constructions in a language as binary properties. The Adj≺N and N≺Adj, as well as the Num≺N and N≺Num orders, are encoded as properties that tell whether a language allows a certain word order. Languages are therefore divided into three types with respect to adjective-noun word order: Adj≺N, N≺Adj, or both word orders are available, such as Romance languages (Table 2). Analogously, there are three types of languages with respect to numeral-noun word order. In total, there are nine possible types of different grammatical word order combinations. Table 2 presents the counts of these language types in the SSWL sample. A new observation is that the counts for rigid word order languages (in light gray) are distributed very similarly to the corresponding counts of languages extracted from WALS and thus re-state Universal 18. The least frequent type is the Adj≺N & N≺Num (one language, in bold), as expected. Furthermore, the counts for flexible word order languages (at least one word order alternates) also indirectly confirm Universal 18. Languages that combine Adj≺N & N≺Num orders (even if it does not dominate) are less frequent than other types. For example, only six languages allow Adj≺N & N≺Adj and N≺Num and only eight allow Adj≺N and Num≺N & N≺Num (in dark grey) while there are 29 and 14 alternating languages that do not allow Adj≺N & N≺Num combination (Adj≺N & N≺Adj with Num≺N and Num≺N & N≺Num with N≺Adj).

Table 2. Number of languages in SSWL database having one of the combinations of word order properties

	Num N, N Num	**Num N, –**	**–, N Num**	**no info**	**total**
Adj N, N Adj	16	29	6	4	55
Adj N, –	8	40	1	2	51
–, N Adj	14	16	48	3	81
no info	3	4	4	14	25
Total	**41**	**89**	**59**	**23**	**212**

In summary, counts collected from typological databases that describe the properties of the word orders of whole languages support the quantitative generalisation expressed in Universal 18, both according to the notion of dominant word order expressed in WALS, and also based on data that take optionality of word orders into account, such as those in SSWL.

Table 3. Percentage of prenominal adjectives and number of alternating adjectives in five Romance languages

Language	nb. instance	% prenominal	nb. alternating adjs
Spanish	20809	27	445
Catalan	17367	22	251
Italian	2868	34	147
Portuguese	7565	30	175
French	15806	27	256

It is important to note, however, that Greenberg's Universal 18 and the WALS counts refer to the dominant word orders of languages.[4] The dominant word order values in the WALS database are given according to the two criteria of unmarkedness and frequency and, in our work, we assume that dominant word order is equivalent to the most frequent word order.

This frequency-based notion hides the fact that word orders can vary. Adjectives in Romance languages are a well-known example. While in these languages adjectives occur predominantly after the noun, there is also considerable variation. Some adjective positioning changes meaning (a 'pauvre femme' is pitiable, but a 'femme pauvre' is a pauper) or restricts the number of meanings available ('une ancienne eglise' can be either a former church or an old church, but a 'eglise ancienne' is only old.). Other adjectives can be only postnominal or only prenominal, but many can alternate based on many other factors. (See Gulordava and Merlo (2015); Gulordava et al. (2015) for corpus-based discussion of the factors of these alternations.) Quantitatively, Table 3 shows the frequencies of five Romance languages, in a sample of 64'000 instances of noun-adjective word orders.[5] The table shows clearly that while the dominant position for adjectives is post-nominal

4. The notion of dominant or basic word order has been largely discussed and challenged in the typological literature (Dryer 1989, 1995; Mithun 1992). Commonly, by basic word order, one understands the unmarked word order, specifically, pragmatically unmarked word order. An exception to this view are the languages that were identified as having discourse-determined word order, where all of possible orders are grammatical and there is no clear neutral word order (Mithun 1992). Yet, in his extensive work, Dryer argues that basic word order can be defined in terms of frequency: the basic word order in a language is that which is used most frequently. See also Haspelmath (2006) who argues that all notions of markedness could be replaced with the notion of frequency.

5. The dependency annotated corpora of five Romance languages are used to collet these counts: Catalan, Spanish, Italian (Hajič et al. 2009), French (McDonald et al., 2013), and Portuguese (Buchholz & Marsi 2006). For more details on the data and data collection, please refer to Gulordava and Merlo (2015).

in all the languages, a sizable number of adjectives occur also prenominally for a total of up to 30% of the token occurrences for some languages. The variation in the positioning of adjectives, then, is not a marginal phenomenon. An even finer-grained investigation, then can be based on syntactically-annotated corpora, studying variation at the sentence level in individual languages, as we show in the next sections.

3. Accounts of language universals and Universal 18

Approaches that try to explain the asymmetries of typological distributions documented in the previous section roughly fall in two categories: the approaches that explain language universals as the result of constraints on language or, more broadly, the cognitive system of native speakers, and the explanations that argue exclusively in favour of a source external to cognition. The latter explanations focus on historical development, spread and contact of languages and speaker communities. For example, a much discussed piece of work by Dunn et al. (2011) models computationally the genetic evolution of four language families and claims to show that word order traits in the evolutionary process were developing independently of each other, whereas a linguistic explanation predicts that they must be correlated. Based on the observation that some traits were dependent during the genetic evolution in one language family but not in all language families, Dunn et al. (2011) conclude that word order universals are culturally dependent and not cognitively universal.

We adopt the competing view which argues that linguistic or cognitive factors play at least some role in shaping typological universals. We present in more detail two families of explanations, one corresponding to a token-level interpretation of word order universal 18 (Cinque 2005; Biberauer et al. 2014) and one corresponding to a type-level interpretation (Culbertson et al., 2012). We will then compare their predictions based on corpus data, to select the model that best accounts for the data.

3.1 Structure-level accounts

Typological word order universals are commonly divided in two types: harmonic universals and implicational universals. Harmonic universals constrain the same types of constructions to appear in the same order, for instance, for a dependent to be always on the left of its head.[6] A much cited example is a correlation between

6. We do not adhere to an explicit definition of the terms 'head' and 'dependent'. Most linguistic theories, both formal and functional, define these terms in some specific way, but for

the order of object and verb and the order of noun and adposition. In generative grammar, these harmonic tendencies are captured by the head-directionality parameter (Baker 2001).

In addition to harmonic universals, we can recognise a second type of universals, referred to as *implicational*, such as Universal 18. Implicational universals are generalisations of robust asymmetries between two disharmonic word order options. Consider the four possible combinations of two orders between H_1 and Dep_1 and H_2 and Dep_2 ($\prec$ means 'precedes'):

$$O1 - H_1 \prec Dep_1 \,\&\, H_2 \prec Dep_2; \quad (1)$$

$$O_2 - Dep_1 \prec H_1 \,\&\, Dep_2 \prec H2; \quad (2)$$

$$O3 - H_1 \prec Dep_1 \,\&\, Dep_2 \prec H_2; \quad (3)$$

$$O_4 - Dep_1 \prec H_1 \,\&\, H_2 \prec Dep_2. \quad (4)$$

Orders O1 and O2 are harmonic, while O3 and O4 are disharmonic combinations. An implicational universal is stated when one disharmonic combination (e.g. O3) is observed much more frequently than another disharmonic combination (O4) across languages. These kinds of universals are not predictable under the head-directionality parameter. A recent proposal is to appeal to the Final-over-Final condition (FOFC) (Biberauer et al. 2014), introduced by Holmberg (2000). The FOFC states that syntactic structures where a head-final phrase ($Dep_1 \prec H_1$) takes as its complement a head-initial phrase ($Dep_1 = H_2$, $H_2 \prec Dep_2$) are prohibited in syntax. Consider Figure 1, which illustrates four possible structures between the phrase αP with the head α and the phrase βP with the head β which takes αP as its complement. These structures correspond to the four possible word order combinations O1–O4. FOFC bans the disharmonic structure (Figure 1, example d) creating therefore an asymmetric account of the two disharmonic structures (Figure 1, example c) and (Figure 1, example d).

The generalisation discussed in Biberauer et al. (2014) can in principle apply to the asymmetry captured by Universal 18.[7] This explanation depends on the specific structure that is assumed for the noun phrase. It presupposes that the structure of the noun phrase has an additional functional head X^0, associated with a particular adjectival position, and that this position takes the noun as its

---

most of the discussion we will assume a simple interpretation of the head-dependent (as well as the head-modifier) relation: we take noun to be the head of noun phrase and adjectives and numerals to be modifiers of the noun; verb to be the head of verb phrase and object to be its complement; adposition to be the head and take a noun as its complement.

7. Culbertson et al. (2012), which we discuss later, mention FOFC as a possible explanation for the suppression of Adj≺N & N≺Num order with respect to another disharmonic order Num≺N & N≺Adj.

complement to form an XP phrase. The numeral, in turn, takes XP as its complement to create a NumP. If this is the case, then the Adj≺N≺Num order must be the linearisation of a structure (with or without movement) of type [[Adj N] Num], directly violating FOFC.[8]

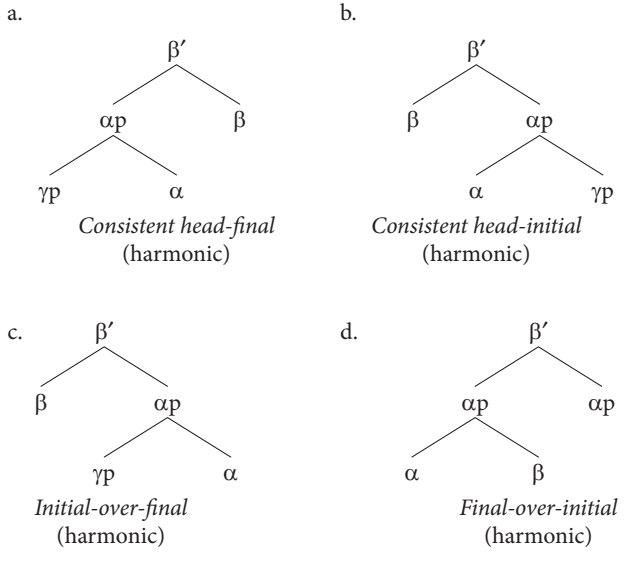


Figure 1. Four possible complementation combinations of two phrases *α*P and *β*P with respect to their head-directionality replicated from Biberauer et al. (2014, p. 117)

8. The assumption of the hierarchical structure [[Adj N] Num] for the noun phrase is controversial since, in many syntactic approaches, the adjective phrase is treated as an adjunct to the noun (there could be several adjective phrases modifying the same noun), and not a higher projection which takes a noun as its complement. Interestingly, the alternative of treating adjectives as lower dependents of N was employed to explain a different set of word order universals. The Head-final Filter of Williams (1982) was proposed as a constraint against [[*α γ*P] *β*] structures arising in noun phrases such as **[proud [of his son]] man*. Sheehan et al. (2017) discusses the close relation between FOFC and the Head-final Filter. The main motivation for the Num > Adj > N hierarchical structure comes from the notion of semantic scope (Rijkhoff, 1998). Adjectives typically modify the internal properties of the noun, while numerals do not. In English, while both Adj≺Num≺N ("heavy three boxes") and Num≺Adj≺N ("three heavy boxes") are possible, the first option has a special interpretation with the adjective having its scope over the numeral: the three boxes are heavy together (Ouwayda 2014; Schwarzschild 2009). The iconicity principle suggests that the semantic hierarchy Num > Adj > N should be mapped onto the syntactic structure (Rijkhoff 1998; Culbertson & Adger 2014).

FOFC is put forward as a categorical constraint to capture a large set of syntactic phenomena cross-linguistically. The strong categorical assumption of FOFC runs into empirical problems when confronted with the gradual nature of constraints observed in typological distributions and language variation. Hawkins (1994), for example, suggested a processing constraint targeting [[α γP] β], involving memory or computational constraints, while Abels (2013) proposes that this structure is disfavored because of its misalignment with prosodic boundaries. For our purposes, the important common part of these accounts is that they all postulate a (hard or soft) constraint against a particular structural realisation of the complete NP phrase.

3.2 Grammar-level accounts

Experimental work by Culbertson et al. (2012) gives a type-level account of Greenberg's Universal 18. It takes as a starting point the original formulation by Greenberg involving the dominant adjective-noun and numeral-noun word order properties of a language. The results of an artificial learning experiment where participants learn a language of one of four possible word order types L1–L4 show, they argue, that the markedness of Adj$\prec$N & N$\prec$Num type is a result of a cognitive learning bias.

The experiment by Culbertson et al. (2012) consists in learning a small artificial language with a grammar defined by two probabilistic phrasal rules, which specify a dominant combination of adjective-noun and numeral-noun word orders. Specifically, participants receive as learning input instances of noun phrases of two types: adjective plus noun phrases, i.e. the artificial language equivalent of "green apple", or numeral plus noun phrase, i.e. the equivalent of "three apples". The lexicon is taught to participants using a visual interface. Four groups of participants are taught four different types of languages: each language has a grammar corresponding to one of four possible dominant word order combinations L1–L4. These grammars G_1–G_4 are presented in Table 4.

Table 4. Four types of grammars taught in the artificial language learning experiment of Culbertson et al. (2012)

	Adj N	**N Adj**
Num N	$G_1 = G_{0.7,0.7}$	$G_2 = G_{0.3,0.7}$
N Num	$G_3 = G_{0.7,0.3}$	$G_4 = G_{0.3,0.3}$

Each grammar G_i is probabilistic, with parameters α and β corresponding to the percentages of instances of dominant adjective-noun and numeral-noun orders which participants receive as an input, as in the following:

$$G_{\alpha,\beta} = \left\{ \begin{array}{ll} S_{adj} \xrightarrow{\alpha} \text{Adj N} & S_{num} \xrightarrow{\beta} \text{Num N} \\ S_{adj} \xrightarrow{1-\alpha} \text{N Adj} & S_{num} \xrightarrow{1-\beta} \text{N Num} \end{array} \right\}. \quad (5)$$

For example, a participant that learns G_1 will see the dominant Adj≺N word order 70% of the time and the alternative N≺Adj order 30% of the time, and dominant Num≺N order 70% of the time together with N≺Num 30% of the time. After the learning phase, a participant is asked to produce adjective phrases and numeral phrases describing the pictures of objects.

By introducing learners to input with variation, this mixture-shift paradigm allows the experimenter to observe the extent to which the learners alter the language grammar to bring it in line with hypothesised learning biases. Culbertson and colleagues hypothesise several biases which predict different learning outputs. First of all, following many psycholinguistic studies in language learning, they hypothesise a regularisation bias – a general learning bias which constrains the acquired grammar to minimise its variation. This bias predicts that learners will alter a grammar with variation rate (α,β) to a grammar where word order between a modifier and a noun is more fixed, a grammar $G_{\alpha',\beta'}$ where $|\alpha' - 0.5| > |\alpha - 0.5|$ and $|\beta' - 0.5| > |\beta - 0.5|$. For example, if the input grammar has $\alpha = 0.7$, in production, the dominant Adj≺N word order will be used more than 70% of the time ($\alpha' > 0.7$) and if $\alpha = 0.3$, the dominant N≺Adj order will be used more than 70% of time ($\alpha' < 0.3$).

Another type of bias considered in the study is a substantive bias – a bias that drives learners away from some particular structure. Culbertson and colleagues's central hypothesis is that the typological Universal 18 is indeed a reflex of a substantive bias against the Adj≺N & N≺Num word order combination. If this is the case, then the learners' output will reflect this dispreference by showing a different behaviour from standard regularisation behaviour, predicted by the general regularisation bias. The findings of the study confirm these predictions. While participants that were learning languages with grammars G_1, G_3, G_4 showed regularizing behaviour (average use of majority order was significantly greater than 70%), the group that was learning the language of the "dispreferred" type with grammar G_2 did not (its average use of the majority order was below 70%). Both a mixed-effect analysis and Bayesian modeling confirm statistically that these experimental data are best explained by postulating a learning bias against the Adj≺N & N≺Num word order combinations. The conclusion of the authors is that a substantive learning bias against a particular combination of word orders does exist and plays a role in shaping typological distribution of word orders. Moreover, they argue that the fact that it was observed during the short-time of a learning experiment

rules out accounts of word order universals that place the origins of typological asymmetries in factors external to cognition.

Very importantly, a Bayesian model is used to provide an explicit account of Universal 18 (Culbertson & Smolensky 2012). This account assumes, following the experimental design, that the learning bias affects two word order properties of the grammar under study, represented by parameters α and β. These parameters are independent in the grammar, and their joint distribution is only affected by the substantial bias against Adj≺N & N≺Num combination. Specifically, they model the process of acquisition as learning the values of the two parameters α and β. In this model, the bias is represented as a prior probability on the α and β values. Universal 18 is therefore accounted for in terms of the probabilistic constraint on two type-level properties.

4. Our approach to Universal 18

The studies presented in the previous section argue for the existence either of a structural linguistic constraint or of a cognitive language learning bias to explain the cross-linguistic dispreference for the Adj≺N & N≺Num word order combination. We provide further evidence in favour of this view by studying quantitative patterns of NP-internal word order variation across languages in corpora.

We follow here the line of work which connects *quantitative* patterns of syntactic variation to underlying grammatical properties (see for example Merlo and Stevenson (2001); Bresnan et al. (2001); Merlo and Stevenson (2004); Bresnan et al. (2007); Hawkins (2009); Samardžić and Merlo (2012, 2018)). Frequency is known to be the product of many factors external to the grammar, such as the topic of discourse, the style of a speech or a text, and social conventions. However, these pieces of work convincingly show that at least in some cases frequency does reflect grammatical properties. Our investigation of the accounts of Universal 18 applies this same reasoning. We group the accounts of Universal 18 reviewed in the previous section into two types, according to their basic underlying assumptions: structural, token-level accounts, and grammar-level, type-level accounts.

The first group includes, e.g., Biberauer et al. (2014), Svenonius (2007) and Cinque (2005). All these accounts assume that the constraint blocks out some particular structure of the noun phrase, for example the [[Adj N] Num] structure. It is the presence of all of the elements – adjective, numeral, noun – and their arrangement in this particular structure which triggers the constraint. A noun phrase without a numeral will not meet this structural condition and will not be affected

by a structural constraint. We can say, therefore, that the constraint operates at the token level as it restricts only certain realisations of a noun phrase.

The second type of accounts include Culbertson et al. (2012), and also Cysouw (2010), and corresponds to the grammar-level interpretation of the Universal 18. When discussing this work, we underlined that their probabilistic Bayesian account assumes a bias against a combination of two grammar-level surface properties and not against a particular noun phrase structure. The two word order properties were defined by parameters $\alpha = p(\text{Adj} \prec \text{N})$ and $\beta = p(\text{Num} \prec \text{N})$. A grammar is assumed to have one of four possible combinations of these properties and a learning bias is assumed to restrict grammars that have $\alpha > 0.5$ and $\beta < 0.5$. Therefore, we call this account a type-level account, as the bias operates on aggregated word order properties of a language and does not affect every instance of a noun-phrase directly.

These two types of accounts differ substantially in their treatment and interpretation of Universal 18, and provide different predictions for the word order variation patterns within a language. We formalise the difference between the structural and the grammar-level accounts in terms of probability distributions of word orders. Two basic assumptions underlie our account: (i) the dominant word order corresponds to the most probable word order; (ii) the effect of a constraint or bias is reflected in the probability distribution of word order alternatives. The first assumption is largely motivated by the typological work on basic word order discussed above. The second assumption reflects our view that grammatical properties induce frequency effects in corpora of language use.[9]

We start by defining two contexts, C_1 and C_2, that distinguish between the two types of accounts.

1. C_1 when adjectives and numerals occur independently in different noun phrases: C_1 = {noun phrases: adjective + noun = "old + book", or noun phrases: numeral + noun = "three + books"}
2. C_2 when adjectives and numerals co-occur in a noun phrase: C_2 = {noun phrases: adjective + numeral + noun = "three + old + books" }

9. This view is also articulated in Bresnan et al. (2001), where it is argued that grammatical constraints that apply categorically in certain languages, also show up as soft constraints in other languages. Universal 18 appears as a categorical restriction in certain languages, yielding the typological distribution discussed in Section 2, and it shows up as a more gradual constraint on frequencies in other languages, such as Latin and Ancient Greek. We will express this notion of gradation as probability distributions below, although this is not the only way of expressing gradation.

Following the notation from the previous sections, we define the parameters of word order distributions to be $\alpha = pc_1(\text{Adj} \prec \text{N})$ and $\beta = pc_1(\text{Num} \prec \text{N})$. These are word order probabilities for the context C_1. We define the four-way distribution of possible word orders in context C_2 as $\boldsymbol{\gamma}$. If the two contexts C_1 and C_2 are equivalent, then $\boldsymbol{\gamma}$ should be equal to a joint distribution of the two word order distributions $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ defined by two scalar parameters α and β:

$$\boldsymbol{\gamma}(x, y) = \boldsymbol{\alpha}(x) \cdot \boldsymbol{\beta}(y) \tag{6}$$

(where x is the word order between adjective and noun and y is the word order between numeral and noun); and in particular:

$$\gamma(\text{Adj} \prec \text{N}, \text{Num} \prec \text{N}) = \alpha \cdot \beta \tag{7}$$

Then $\boldsymbol{\alpha}(x) \cdot \boldsymbol{\beta}(y)$ is the probability (and the expected frequency) of word orders in C_2 under the hypothesis that the two word orders are independent and the two contexts are equivalent.

Structural accounts predict that the constraint underlying Universal 18 applies in context C_2, but not in context C_1. We should therefore observe an effect of this constraint as a categorical or statistical difference between word order distributions of context C_1 and context C_2. Specifically, in case of statistical bias, structural accounts predict that the observed frequency of the Adj≺N≺Num word order in the context C_2 must be smaller than the expected frequency (under the assumption that Adj≺N and N≺Num phrases could freely combine), that is:

$$\boldsymbol{\gamma}(\text{Adj} \prec \text{N} \prec \text{Num}) < \boldsymbol{\alpha}(\text{Adj} \prec \text{N}) \cdot \boldsymbol{\beta}(\text{N} \prec \text{Num}) \tag{8}$$

Type-level accounts offer a different prediction. Specifically, they assume that the bias affects the combination of the two parameters α and β in the language, but not word order combinations in each sentence of a language. These types of accounts do not distinguish therefore between the two contexts C_1 and C_2. Consequently, the two distributions $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha} \cdot \boldsymbol{\beta}$ should be equal.[10]

Given these two different predictions, we can use corpus data to test which theory complies better with it. To observe a possible effect of the structural constraint, we need to look at languages where both Adj≺N and N≺Num are possible grammatical orders (as dominant or alternative orders). While Romance languages show substantial variation in the prenominal or postnominal positioning of adjectives, as discussed above, the numeral almost exclusively precedes the noun. We must look into the predecessor of Romance languages, Latin, to find

10. In principle, C_1 and C_2 could be different and give rise to different word order distributions, however, it is important for us here that this difference is not due to the bias against Adj≺N≺Num.

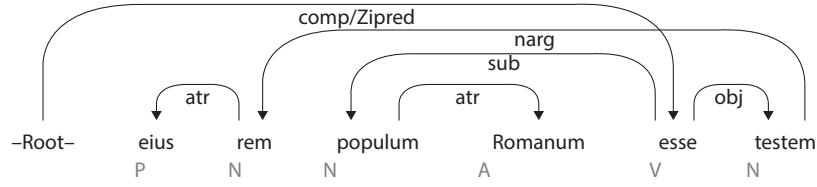
substantial variation of both these orders. The languages in our sample that show the required variations are Latin and Ancient Greek. We quantitatively analyse the corpora of these languages to confirm the predictions on observed data. Specifically, we collect counts corresponding to α, β, γ. Using these counts, we test the null hypothesis that observed counts γ correspond to expected counts sampled from the distribution $\alpha \cdot \beta$. Accepting the null hypothesis will give evidence in favour of grammar-level accounts; rejecting the null hypothesis will provide evidence in favour of structural accounts. In the next sections, we describe our data, the procedure to collect the counts and present our results.

4.1 Cross-linguistic corpus data

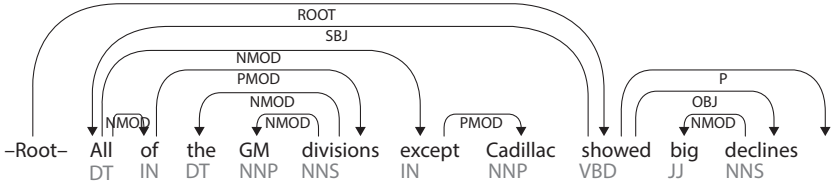
Corpus data constitute important evidence for studying syntactic variation. To induce reliable generalisations from corpora, one needs to ensure that the sample of extracted constructions is both accurate and representative of the language's grammar and variation. We use gold syntactically annotated corpora, since we analyse certain structural constructions in a language and not just collocations of words. A drawback of using gold-annotated corpora is their limited size since the manual syntactic annotation is expensive to produce. Most of the recent corpora come in dependency format. The features necessary for our task include part-of-speech (PoS) tags of the words, the head of the current token (the dependent of the head), and the type of dependency relation between the token and the head. These basic features are illustrated in Figure 2, for two typologically different languages. For our investigation, we extract all noun phrases that contain either adjective or numeral modifiers (or both) from our corpora.

Corpus data for Latin and Ancient Greek

The corpora for Latin and Ancient Greek come from the Perseus project (Bamman & Crane 2011) and from the PROIEL project (Haug & Jøhndal 2008). The Latin Perseus corpus is small, and in our study of the co-occurrences of adjectives and numerals in a noun phrase, we observed only 16 such cases. The Ancient Greek Perseus corpus consists for the most part of poetic works (including Homer's Iliad and Odyssey, Aeschylus' Agamemnon and others). Rhythmic constraints in poetic work significantly affect the choice of word order and it could therefore influence our results. The PROIEL Ancient Greek corpus contains exclusively prose and is therefore more appropriate for a word order variation study. We therefore base our analyses and conclusions primary on the PROIEL sources. The Perseus treebanks remain nevertheless important additional data which corroborate our results.



a. Sample sentence in Latin



b. Sample sentence in English

Figure 2. Illustration of dependency annotation, visualised as dependency arcs and their functional labels

Preprocessing and collection of counts

We preprocess the PROIEL corpora to obtain the universal PoS tags (Petrov et al. 2012) and dependencies as our basic annotation information. The original annotation of the PROIEL corpora distinguishes between cardinal and ordinal numerals. The case of Universal 18 concerns only cardinal numerals, which differ in their structural and functional properties from ordinal numerals. The preprocessing ensures that only numerals tagged as cardinals in the original corpora are included in the analysis.

We collect two types of counts for our analysis. First of all, to estimate the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ distributions, we extract the frequencies of pre- and post-nominal adjective-noun and numeral-noun word orders. In addition, to estimate the $\boldsymbol{\gamma}$ distribution, we extract the same pair-wise word order frequencies in the restricted context of noun phrases that contain both an adjective and a numeral. The total counts for word order combinations are then computed from the output.

Collected counts

Table 5 displays raw counts and relative frequencies – α and β parameters – for adjective-noun and numeral-noun word order pairs collected from the Ancient Greek and Latin PROIEL corpora. We present frequencies for each literary work separately.

Table 5. The percentages of pre-nominal and post-nominal placement of numerals and adjectives in the Ancient Greek and Latin PROIEL corpus. Parameter α is estimated as the percentage of Adj≺N and β is estimated as the percentage of Num≺N. Abbreviations: Chr. = Chroniche; NT = New Testament; Her. = Herodotus; Per. = Peregrinatio; Vul. = Vulgate

	Chr.	**NT**	**Her.**	**Total**	**Caesar**	**Cicero**	**Per.**	**Vul.**	**Total**
Adj≺N	285	648	768	1701	614	629	320	412	1975
N≺Adj	111	1042	978	2131	238	462	272	794	1766
α	**72.0**	**38.3**	**44.0**	**44.4**	**72.1**	**57.7**	**54.1**	**34.2**	**52.8**
Num≺N	26	287	215	528	88	71	65	257	481
N≺Num	39	146	101	286	58	26	18	118	220
β	**40.0**	**66.3**	**78.0**	**64.9**	**60.3**	**73.2**	**78.3**	**68.5**	**68.6**

Table 6 displays observed co-occurrence counts for adjective-noun and numeral-noun word orders – the counts used to estimate the γ distribution in context C_2. The tables also include the respective *expected* co-occurrence counts: the counts that would be produced under the joint distribution $\alpha \cdot \beta$ (α and β are estimated from Table 5) given that the context C_2 of a noun phrase containing both an adjective and a numeral was not different from the general context C_1. Each expected count for word orders x and y is calculated proportionally to $\alpha(x) \cdot \beta(y)$ and the observed total number of noun phrases where two modifiers co-occur. We compare the observed and the expected counts to draw the generalisations about constraints on adjective and numeral syntactic distributions.

Table 6. Co-occurrence counts in the Ancient Greek and Latin PROIEL corpus, divided by source text. Abbreviations: Chr. = Chroniche; NT = New Testament; Her. = Herodotus; Cae. = Caesar; Cic = Cicero; Per. = Peregrinatio; Vul.=Vulgate.

		Chr.	**NT**	**Her.**	**Tot**	**Cae.**	**Cic.**	**Per.**	**Vul.**	**Tot**
					Observed counts, γ					
Adj≺N	Num≺N	1	10	3	14	5	3	2	2	12
Adj≺N	**N≺Num**	**0**	**2**	**2**	**4**	**1**	**1**	**0**	**2**	**4**
N≺Adj	N≺Num	1	3	13	17	7	4	3	3	17
N≺Adj	Num≺N	1	9	15	25	1	5	12	15	33
	Total	3	24	33	60	14	13	17	22	66
					Expected counts, $\alpha \cdot \beta$					
Adj≺N	Num≺N	0.9	6.1	9.9	17.3	6.1	5.5	7.2	5.2	23.9
Adj≺N	**N≺Num**	**1.3**	**3.1**	**4.6**	**9.4**	**4.0**	**2.0**	**2.0**	**2.4**	**10.9**
N≺Adj	N≺Num	0.5	5.0	5.9	11.7	1.6	1.5	1.7	4.6	9.8
N≺Adj	Num≺N	0.3	9.8	12.6	21.6	2.4	4.0	6.1	9.9	21.4

4.2 Results and discussion

Overall there were only 60 cases for Ancient Greek and 66 for Latin of noun phrases with both an adjective and a cardinal numeral modifier. If we take each work separately, the counts are insufficient to allow statistically significant tests. However, combining observations for multiple texts, and especially for the two languages together, we can perform a relevant cumulative statistical analysis. An appropriate analysis would include a multilevel statistical model, such as a mixed-effect model, with word order distribution parameters varying for each text and a bias parameter equal for all texts and languages. Unfortunately, this approach is not applicable given the small number of observations for individual texts. We assume that the cumulative χ^2 statistical tests we present here provide a sufficient approximation to the actual levels of statistical significance.

The main observation arising from the data in Table 6 is that the word order combination Adj≺N≺Num is always underobserved with respect to the expected counts. It is true for both languages and for all texts, despite the few counts that are available for analysis. Cumulatively, while the observed four-way distribution for Ancient Greek is not statistically different from the expected distribution ($\chi^2 = 6.6$, $p = 0.085$), the difference between counts is significant for Latin ($\chi^2 = 22$, $p < 0.001$). Most importantly, we evaluate a possible bias against the Adj≺N≺Num word order by comparing statistically expected and observed two-way distributions: Adj≺N≺Num word order versus all other word order combinations. For Ancient Greek we therefore compute the χ^2 test on pairs of counts: observed counts (4; 56) versus expected counts (9.4; 50.6), which is marginally significant ($\chi^2 = 3.6$, $p = 0.056$). For Latin, the difference between observed counts and expected counts, (4; 62) v (10.9; 61.1), is also significant ($\chi^2 = 5.3$, $p = 0.02$). Furthermore, combining counts for the two languages and testing the two-way distribution yields a significance level of $p = 0.003$. Despite the limited number of co-occurrences of adjectives and numerals, our data gives a very interesting and at least marginally significant observation: the Adj≺N≺Num word order, corresponding to the dispreferred order in Universal 18, is under-observed in all cases.

Note that counts for other word order options do not show a similarly constant pattern. In Latin, Adj≺N & Num≺N order is also significantly under-observed, but this does not hold for Ancient Greek. We also controlled for the size of numeral and adjective phrases as a possible explanatory factor in word order variation. The distribution of simple and complex (longer than one word) adjective phrases and numeral phrases were the same for both samples: the sample of pair-wise word orders and the sample of word orders in co-occurrence. This factor therefore could not be the reason for the persistent differences in two distributions of word orders.

On the basis of the statistical tests and observations above, we therefore reject the null hypothesis corresponding to the prediction of the grammar-level

accounts. The distributions for modifiers on their own and together differ in a way which confirms that they are not independent and, moreover, that there is a bias against one particular word order combination: we have observed that when adjectives and numerals co-occur there is a statistical tendency to skew the distributions away from the typologically dispreferred Adj≺N≺Num word order.

Our final conclusion is two-fold: first, the word order variation data support the accounts of Universal 18 that argue for its language-internal origin; secondly, our data speaks in favour of the token-level accounts and against the type-level accounts. Specifically, the data points to a linguistic bias which constrains token-level distributions and not type-level properties. Recall that, although Universal 18 is stated typologically for dominant word orders, we have observed, based on the data in SSWL, that its structural interpretation is also confirmed for a number of languages. Therefore, there seems to be converging evidence in favour of structural accounts.

Nevertheless, this view faces problems when attempting to explain both the experimental results of Culbertson et al. (2012) and our corpus data under a single theory. Culbertson et al. (2012) mention FOFC as a potential explanation for Universal 18. However, some additional considerations are necessary to unify type-level and token-level accounts. In the learning experiment, participants received input adjective phrases and numeral phrases independently. They never observed the realisations of the dispreferred Adj≺N≺Num order. The learners, however, still showed a bias against the language type L2 (Adj≺N & N≺Num). To claim that this is expected under FOFC or any other structural account, one needs to specify how the learners induced the underlying grammar and particularly, how they induced the structure of the complete noun phrase from the input data they were given. One possibility is to say that the complete NP structure is entertained even if only a partial overt realisation (e.g., adjective plus noun) is observed. Such an assumption is in line with frequent adherence to empty projections in generative syntax and an extended NP phrase structure (with different adjective functional projections). If this is the case, then this explains the results of Culbertson et al. (2012), but we still need to answer why in our corpus data there is an additional bias observed in noun phrases with both an adjective and a numeral compared to distributions for partial NPs. We could speculate that a structural constraint is triggered when all the projections are overt.

5. Towards a model explaining Universal 18

In this section, we sketch the reasoning needed to develop an explanatory model of the observed corpus distributions of nominal modifiers for Latin and Ancient Greek. We would like to statistically compare a model which incorporates a bias

predicted by the Universal 18 and a model without such bias. Following the discussion in the previous sections, crucially, our problem includes two types of data – adjective-noun and numeral-noun occurrences (call these data D_1) and co-occurrences of adjectives and numerals inside a single noun phrase (call these data D_2). Given these data of two distinct but related types, we define the grammatical phenomenon (or process) we want to model as what is expressed in D_2 and Universal 18. In particular, our abstract models must specify what their parameters are, how the parameters are estimated from the data, and which data are used to test the competing models.

The third question receives a natural answer right away, as it is D_2 which shows the biased distribution of modifiers and, therefore, is the natural choice for testing the significance of the bias effect. In the following, we will try to narrow down the possible answers for the first two questions.

The first two intuitive parameters that we assume at the moment are α and β, which are the probabilities for the Adj≺N and Num≺N orders. These are also the parameters that Culbertson and colleagues use in their experiment and its Bayesian modelling. A word order model that has only these two parameters (we can call it $G(\alpha,\beta)$) is a minimal model. Figure 3(a) (Model 1) depicts this minimal model with two parameters α and β. Importantly, this model assumes that D_1 and D_2 are generated by the same grammar. This also means that we can choose to estimate the model parameters from D_1, D_2 or both.

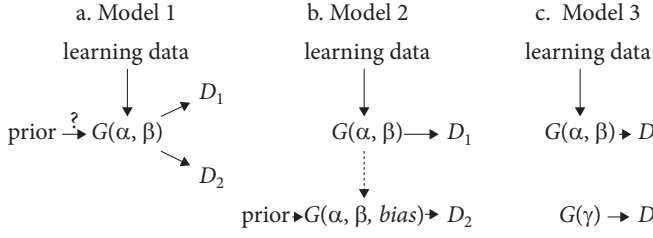


Figure 3. Three models applicable to the data D1 and D2

On the opposite side of the modelling spectrum, we can devise a maximal model (Figure 3(c), Model 3). This model assumes that the two types of data are generated by two different grammars. While D_1 is generated by $G(\alpha,\beta)$ (which is a maximal model for D_1 in the sense that it uses two parameters to explain two independent binary distributions), D_2 is generated by a four-way distribution γ. The parameters are straight-forwardly estimated from D_1 and D_2 correspondingly. Model 3 has a higher number of free parameters than Model 1, and therefore can fit the data better. For our data set the resulting log-likelihood of Model 3 was equal to −24.13. However, this model does not reflect our intuition that $G(\alpha,\beta)$ and $G(\gamma)$ are related

and does not really correspond to any existing theory. Even usage-based theories would predict, we think, a link between α and β and γ, possibly with a data dependency from γ to α and β. For this reason we do not further discuss Model 3.

From this point of view, the sketch of Model 2 (Figure 3(b)) presents a more satisfactory model. The property of such model would be the coupling of the parameters of $G(\alpha,\beta)$ and $G(\gamma)$. In particular, we can assume that $G(\gamma)$ is a combination of $G(\alpha,\beta)$ and a structural bias which could be modelled with as few as one parameter. While the α and β parameters can be estimated from D_1, the bias parameter(s) must be necessarily estimated from D_2.

Formalized in this way, the models are easily compared to each other. Comparison of Model 1 and Model 2 can answer the question whether there is a significant bias effect in the co-occurrence of modifiers.

5.1 Comparison of models 1 and 2

We assumed that both models are probability distributions (corresponding to probabilities of four possible word order combinations: $p1 + p2 + p3 + p4 = 1$). We computed the likelihoods of the two models on D_2 data given their parameters estimated from D_1 data for Model 1 and both D_1 and D_2 data for Model 2.

Specifically, in both cases we calculate the likelihoood as a multinomial dependent on the four parameters, ($C(\cdot)$ indicates frequency counts):

$$p(D_2 \mid \text{Model 1}) = Mult\left(C(D_2) \mid p_1, p_2, p_3, p_4\right) \tag{9}$$

For Model 1, the probabilities of the distribution come from the $\alpha\cdot\beta$ joint probabilities:

$$p_1 = \alpha\cdot\beta;\ p_2 = (1-\alpha)\cdot\beta;\ p_3 = \alpha\cdot(1-\beta);\ p_4 = (1-\alpha)\cdot(1-\beta). \tag{10}$$

For Model 2, we introduce one more parameter for the bias as follows, where b is a normalisation parameter, so that $p1 + p2 + p3 + p4 = 1$:

$$\begin{aligned} p_3 &= p(\text{Adj} \prec \text{N} \ \&\ \text{N} \prec \text{Num}) = \alpha\cdot(1-\beta)\cdot bias \\ p_1 &= \alpha\cdot\beta\cdot b \\ p_2 &= (1-\alpha)\cdot\beta\cdot b \\ p_4 &= (1-\alpha)\cdot(1-\beta)\cdot b \end{aligned} \tag{11}$$

The bias parameter was estimated on D_2 data using MLE estimation. In other words, the likelihood of Model 2, $p(D_2|Model2)$, is a function of the parameter *bias* and we choose the value of this parameter which maximises the likelihood:

$$bias^{\star} = arg \max_{bias\in[0,1]} Mult_{bias}(D_2 \mid p_1, p_2, p_3, p_4). \tag{12}$$

Note that *bias* is defined to be less than 1, which corresponds to a decreased probability of the dispreferred word order Adj≺N & N≺Num with respect to Model 1.

Model 2 is therefore an extension of Model 1 with one added parameter. The resulting log-likelihoods on D_2 for the two models were log-Likelihood(Model 1) = −55.31 and log-Likelihood(Model 2) = −50.37. We use these log-likelihood values to test the fit of the models. Note that Model 2 has a higher number of free parameters than Model 1, and therefore can fit the data better. Model 2 uses two free parameters per text, overall fourteen parameters, plus one global bias parameter, resulting in fifteen parameters, while Model 1 has the same number of free parameters but no bias, that is fourteen for our set of seven texts. According to a χ^2-squared test, a test that takes into account the degrees of freedom, the difference between the fit of Model 1 and the better fit of Model 2 is statistically significant at $p < 0.01$.

6. Conclusions

Our results, combined with previous theoretical proposals and experimental findings, can provide new evidence for the debate about language universals, in particualr for the competing explanations provided by generative syntax and usage-based theories of language. The experimental results can be seen as evidence that learners acquire complete noun-phrase structure (including the hierarchy of relevant adjectival and numeral heads) without seeing actual realisations of them. This conclusion is not surprising under current assumptions in generative syntax, which postulate the innate knowledge of basic grammatical structure. Our corpus counts show a clear numerical under-representation corresponding to Universal 18, thereby corroborating this hypothesis. Preliminary definition and testing of explanatory models show that a model where the grammar-level parameters are modulated by a structure-specific bias is more strongly supported by the evidence than a model without bias.

The study of Universal 18 demonstrates that quantitative corpus-based approaches to syntactic variation can inform formal syntactic approaches and in particular can provide more precise specification of the models, thereby providing the means for precise models selection. We showed that by studying quantitative patterns of NP-internal word order variation in Latin and Ancient Greek we can provide new evidence that discriminates between two types of theoretical accounts, thereby teasing apart fine-grained difference in proposals about a typological universal by looking at language-internal word order variation.

References

Abels, Klaus. 2013. "On 2-3-1." *UCL Working Papers in Linguistics* 25.

Baker, Mark. 2001. *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York, NY: Basic Books.

Bamman, David, and Gregory R. Crane. 2011. "The Ancient Greek and Latin Dependency Treebanks." In *Language Technology for Cultural Heritage* (Theory and Applications of Natural Language Processing), ed. by Caroline Sporleder, Antal Bosch, and Kalliopi Zervanou, 79–98. Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-642-20227-8_5

Biberauer, Theresa Anders Holmberg, and Ian Roberts. 2014. "A Syntactic Universal and its Consequences". *Linguistic Inquiry* 45 (2): 169–225. https://doi.org/10.1162/LING_a_00153

Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and Harald Baayen. 2007. "Predicting the Dative Alternation." In *Cognitive Foundations of Interpretation*, ed. by G. Boume, I. Kraemer, and J. Zwarts, 69–94. Amsterdam: Royal Netherlands Academy of Science.

Bresnan, Joan, Shipra Dingare, and Christopher D. Manning. 2001. "Soft Constraints Mirror Hard Constraints: Voice and Person in English and Lummi." In *Proceedings of the LFG 01 Conference*, 13–32. CSLI.

Buchholz, Sabine, and Erwin Marsi. 2006. "CoNLL-X Shared Task on Multilingual Dependency Parsing." In *Proceedings of the Tenth Conference on Computational Natural Language Learning, CoNLL-X '06*, 149–164. Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=1596276.1596305. https://doi.org/10.3115/1596276.1596305

Cinque, Guglielmo. 2005. "Deriving Greenberg's Universal 20 and its Exceptions". *Linguistic Inquiry* 36 (3): 315–332. https://doi.org/10.1162/0024389054396917

Culbertson, Jennifer, and David Adger. 2014. "Language Learners Privilege Structured Meaning Over Surface Frequency". *Proceedings of the National Academy of Sciences of the United States of America* 111 (16): 5842–5847. https://doi.org/10.1073/pnas.1320525111

Culbertson, Jennifer, and Paul Smolensky. 2012. "A Bayesian Model of Biases in Artificial Language Learning: The Case of a Word-Order Universal". *Cognitive Science* 36 (8): 1468–1498. https://doi.org/10.1111/j.1551-6709.2012.01264.x

Culbertson, Jennifer, Paul Smolensky, and Géraldine Legendre. 2012. "Learning Biases Predict a Word Order Universal". *Cognition* 122 (3): 306–329. https://doi.org/10.1016/j.cognition.2011.10.017

Cysouw, Michael. 2010. "Dealing with Diversity: Towards an Explanation of NP Word Order Frequencies". *Linguistic Typology* 14 (2): 253–287. https://doi.org/10.1515/lity.2010.010

Dryer, Matthew S. 1989. "Discourse-Governed Word Order and Word Order Typology". *Belgian Journal of Linguistics* 4: 69–90. https://doi.org/10.1075/bjl.4.05dry

Dryer, Matthew S. 1992. "The Greenbergian Word Order Correlations". *Language* 68: 81–138. https://doi.org/10.1353/lan.1992.0028

Dryer, Matthew S. 1995. "Frequency and Pragmatically Unmarked Word Order." In *Word Order in Discourse*, ed. by Mickey Noonan, and Pamela Downing, 105–135. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.30.06dry

Dryer, Matthew S., and Martin Haspelmath (eds.). 2011. *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. http://wals.info/

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson, and Russell D. Gray. 2011. "Evolved Structure of Language Shows Lineage-Specific Trends in Word-Order Universals". *Nature* 473 (7345): 79–82. https://doi.org/10.1038/nature09923

Greenberg, Joseph H. 1963. "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements." In *Universals of Language*, ed. by Joseph H. Greenberg, 73–113. MIT Press.

Gulordava, Kristina, and Paola Merlo. 2015. "Structural and Lexical Factors in Adjective Placement in Complex Noun Phrases across Romance Languages". In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, 247–257. Beijing, China.

Gulordava, Kristina, Paola Merlo, and Benoit Crabbé. 2015. "Dependency Length Minimisation Effects in Short Spans: A Large-Scale Analysis of Adjective Placement in Complex Noun Phrases." In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Volume 2: Short Papers), 477–482. Beijing, China.

Hajič, Jan, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. "The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages." In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task, CoNLL '09*, 1–18. Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=1596409.1596411

Haspelmath, Martin. 2006. "Against Markedness (and What to Replace it With)". *Journal of Linguistics* 42 (01): 25–70. https://doi.org/10.1017/S0022226705003683

Haug, Dag T. T., and Marius L. Jøhndal. 2008. "Creating a Parallel Treebank of the Old Indo-European Bible Translations." In *Proceedings of the 2nd Workshop on Language Technology for Cultural Heritage Data*, 27–34. Marrakech, Morocco.

Hawkins, John A. 1983. *Word Order Universals*. New York: Academic Press.

Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

Hawkins, John A. 2009. "Language Universals and the Performance-Grammar Correspondence Hypothesis." In *Language Universals*, Chapter 4, ed. by M. H. Christiansen, C. Collins, and S. Edelman, 54–79. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195305432.003.0004

Holmberg, Anders. 2000. "Deriving OV Order in Finnish." In *The Derivation of VO and OV*, 123–152. Amsterdam: John Benjamins. https://doi.org/10.1075/la.31.06hol

Hurford, James R. 2003. "The Interaction between Numerals and Nouns." In *Noun Phrase Structure in the Languages of Europe*Volume 20–7 of Empirical Approaches to Language Typology, ed. by F. Plank. Mouton de Gruyter.

McDonald, Ryan, Joakim Nivre, Yvonne Quirmbach-Brundage, Yoav Goldberg, Dipanjan Das, Kuzman Ganchev, Keith Hall, Slav Petrov, Hao Zhang, Oscar Täckström, Claudia Bedini, Núria Bertomeu Castelló, and Jungmee Lee. 2013. "Universal dependency annotation for multilingual parsing." In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Volume 2: Short Papers), 92–97. Association for Computational Linguistics. http://aclweb.org/anthology/P13-2017

Merlo, Paola, and Suzanne Stevenson. 2001. "Automatic Verb Classification Based on Statistical Distributions of Argument Structure". *Computational Linguistics* 27 (3): 373–408. https://doi.org/10.1162/089120101317066122

Merlo, Paola, and Suzanne Stevenson. 2004. "Structure and frequency in verb classification." In *Proceedings of Incontro di Grammatica Generativa XXX*, ed. by Laura Brugé, Giuliana Giusti, Nicola Munaro, Walter Schweikert, and Giuseppina Turano, Venice, Italy.

Mithun, Marianne. 1992. "Is Basic Word Order Universal? In *Pragmatics of Word Order Flexibility*, ed. by Doris L. Payne, 16–61. Amsterdam: John Benjamins.

Ouwayda, Sarah. 2014. "Where Number Lies: Plural Marking, Numerals, and the Collective-Distributive Distinction." Ph.D. thesis, University of Southern California.

Petrov, Slav, Dipanjan Das, and Ryan McDonald. 2012. "A Universal Part-of-Speech Tagset." In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, ed. by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, 2089–2096. Istanbul, Turkey, European Language Resources Association (ELRA).

Rijkhoff, Jan. 1998. "Order in the Noun Phrase of the Languages of Europe." In *Constituent Order in the Languages of Europe*, Volume 20–1 of Empirical Approaches to Language Typology, ed. by Anna Siewierska, 321–382. Mouton de Gruyter.

Samardžić, Tanja, and Paola Merlo. 2018. "Probability of External Causation: An Empirical Account of Cross-Linguistic Variation in Lexical Causatives". *Linguistics* 56 (5): 895–938. https://doi.org/10.1515/ling-2018-0001

Samardžić, Tanja, and Paola Merlo. 2012. "The Meaning of Lexical Causatives in Cross-Linguistic Variation". *Linguistic Issues in Language Technology* 7 (1).

Schwarzschild, Roger. 2009. "Stubborn Distributivity, Multiparticipant Nouns and the Count/Mass Distinction." In *Proceedings of the 40th Annual Meeting of the North East Linguistic Society*, Volume 39, 1–18. Mass.: MIT.

Sheehan, Michelle, Theresa Biberauer, Ian Roberts, and Anders Holmberg. 2017. *The Final-Over-Final Condition: A Syntactic Universal*, chapter The Final-Over-Final Condition and the Head-Final Filter. MIT Press. https://doi.org/10.7551/mitpress/8687.001.0001

Svenonius, Peter. 2007. "1…3-2." In *Oxford Handbook of Linguistic Interfaces*, ed. by Gillian Ramchand, and Charles Reiss, 239–288. Oxford University Press.

Williams, Edwin. 1982. "Another Argument that Passive is Transformational". *Linguistic Inquiry* 13: 160–163.

Trajectories of change in Spanish and Portuguese in the Americas

Gregory Guy
New York University

This paper examines ways in which varieties of Spanish and Portuguese spoken in the Americas have diverged significantly from their peninsular sources, and from each other, in the half-millennium since colonization. Some of this divergence is a consequence of spontaneous innovations in the New World varieties (e.g., 'zheismo' and 'sheismo' in Platense Spanish; emergence of the new 1PL pronoun *a gente* in Brazilian Portuguese). Historically, a significant driver of change was language contact, with indigenous languages, and especially with African languages. A suite of linguistic characteristics shared by Caribbean Spanish and Brazilian Portuguese indicate that both varieties were affected by the irregular transmission of these languages to the African population transported to these locations in the time of slavery.

Keywords: language change, language contact, dialects, innovation, African influences

1. Introduction

Spanish and Portuguese are among the handful of languages with a global, multinational reach. This global presence is the result of the colonial conquests set in motion by the 15th century voyages of Columbus and the Portuguese explorers Bartolomeu Dias, Pedro Cabral, and Vasco da Gama. In the Americas, the Spanish *conquistadores* implanted their language over a huge expanse covering two continents, which is today divided among 20 nations. The Portuguese colonial empire covered half of South America, i.e., Brazil, plus an archipelago of territories in Africa and Asia where Portuguese still has official status, but is not the first language of most citizens.

Today, a half millennium later, the varieties of Spanish and Portuguese that are spoken in these far-flung locations are distinctively different from the varieties spoken in Iberia. They are also different from each other: the Spanish of Argentina is appreciably distinct from the Spanish of Mexico, and Brazilian Portuguese shows significant regional dialect differences. This differentiation bespeaks a

https://doi.org/10.1075/rllt.16.09guy

history of language change along different trajectories. Five hundred years is a long time in linguistic history – it is the amount of time that elapsed between Beowulf and Chaucer, or between Chaucer and Robert Frost. More relevant to our case, 500 years is roughly the amount of time that passed between the break-up of the Roman Empire and the differentiation of Latin into the several distinctive Romance languages. While some of the differentiation in Spanish and Portuguese results from changes that occurred in Europe but did not diffuse to the colonies (e.g., change to a stress-timed prosody in Portugal, with consequent reduction or deletion of unstressed segments), my focus here is on the changes that have occurred in the Americas.

In Section 2, I discuss general processes that have driven change in American varieties of Spanish and Portuguese illustrated with some contemporary innovations. Section 3 discusses the remarkable parallel changes that are found in Caribbean Spanish and Brazilian Portuguese, while Section 4 explores possible explanations for these phenomena.

2. Processes of language change in the Americas

As the first languages to undergo global expansion via colonialism and imperialism, Portuguese and Spanish were involved in the entire gamut of drivers of language change. In the Americas, they underwent changes motivated by the following processes: (a) dialect mixture and koinéization, as speakers from different places in the home country came together in settlements in the new colonies; (b) language contact, with indigenous and African languages, and later, in some places, with immigrant populations from Europe and Asia; (c) spontaneous innovations emerging locally in different places, with or without wider diffusion; and (d) standardization, through ongoing contact with the homelands in the colonial era, and via the promotion of national standards by expanded education. The differentiation that we see in the contemporary varieties arises as a consequence of the different circumstances and specific trajectories of change that obtained in each location.

2.1 Dialect mixture

The Portuguese and Spanish spoken in the Americas are "colonial dialects" in Peter Trudgill's (2004) sense, reflecting a dialect mixture among the founding colonizers. In Spanish America, initial colonizers came from different parts of Spain, including primarily speakers of Andalusian and Castilian varieties, and no doubt some of Aragonese, Catalan, Gallego, Basque, etc. In Brazil, the colonizing population included mainland Portuguese, Azoreans, and people coming from

other Portuguese possessions in Africa, such as São Tomé and Angola. Hence the usual questions arise about which groups predominated and how their different dialects influenced or contributed to the emergent colonial dialect. In Spanish American dialectology, for example, there are debates about the relative influence of Castilian and Andalusian dialect features in different regions, and in Brazil, Azorean settlers are presumed to be the source of some regional features in Santa Catarina and Rio Grande do Sul.

2.2 Language contact

Ibero-America saw language contact on a massive scale. There were two main sources of contact languages, indigenous languages of the Americas, and African languages. Indigenous languages, mostly those that Greenberg (1987) calls 'Amerind', were particularly prominent sources of contact in the indigenous empires that the Spanish defeated, notably Aztecan Nahuatl in Mexico, and the Incan Quechua in the Andean region in South America. These languages are still extant with numerous speakers in Peru, Bolivia, Ecuador, Mexico, and contact continues. Mayan languages are still spoken by a substantial fraction of the population of Guatemala and the Mexican Yucatan. Spanish has also had long, intimate, and continuing contact with Guarani in Paraguay. These languages donated loan words to Spanish, and arguably had other contact effects; they are sources of dialect differentiation in the Americas.

In Brazil, Portuguese had sustained contact over two-and-a-half centuries with Tupi-Guarani, leading to a very unusual outcome: the emergence of a widely spoken lingua franca called the *Lingua Geral* 'general language'. This was a Tupi-based koiné, spoken not only by indigenous people, but by Portuguese settlers and explorers on the frontiers of Portuguese colonization. It was so widespread that the Portuguese authorities came to see it as a threat to the dominance of the Portuguese language, leading to the prohibition of its use by the Portuguese crown under the Marques de Pombal in 1757.

The other source of massive language contact was Africa, via the enslaved Africans brought as laborers. Brazil was the foremost destination of the slave trade, receiving some 40% of the entire traffic, approximately 4 million people. Africans were the principal labor force in agriculture (especially sugar cane and coffee cultivation), construction, transportation (porters, stevedores, etc.), and in mining. Brazil was the last country in the Americas to abolish slavery, in 1888.

Hispanic America received another 20–25% of the slave traffic, primarily in the circum-Caribbean territories, namely the Antillean islands (Cuba, Hispaniola, Puerto Rico), and coastal Colombia, Venezuela, Panama. Other populations were taken to Argentina and some points along the Pacific coast, such as Lima. Few

Africans were taken to the interiors of Mexico, Peru, and Bolivia, because the Spanish used indigenous laborers in these places.

Of particular note was a massive increase in the 19th century in the importation of slaves in three countries. In Hispanic America, after the independence movements of the 1810's, the slave trade targeted the two remaining Spanish colonies of Cuba and Puerto Rico, where slavery remained legal until 1866 and 1873. Importation to Brazil also ramped up dramatically around the same time, to the point where people of African descent came to constitute between two-thirds and three-fourths of the Brazilian population by 1851.

The fate of African languages in the Americas was very different from that of indigenous American languages. As we have noted, indigenous languages are still widely spoken in several countries, and the language shift to Spanish or Portuguese took place over many generations. But language shift for most Africans was abrupt: the newly arrived *bozales* were mostly obliged to abandon their native languages – Yoruba, Kimbundu, etc. – and begin to communicate immediately in some version of Portuguese or Spanish. This meant that a substantial fraction of the founding populations in the territories mentioned above were adult Africans undergoing abrupt second language acquisition under the most adverse possible conditions. I will argue that this, perhaps more than any of the other factors mentioned, has had the greatest linguistic impact on the colloquial language of Brazil, Cuba, and the circum-Caribbean generally.

2.3 Spontaneous changes in Latin American languages

The third source of language change in Latin America is a universal one, occurring in all languages at all times, namely spontaneous innovation. The varieties of Spanish and Portuguese spoken in the Americas are currently undergoing a number of documented changes, driving increasing dialect differentiation.

One impressive development is the ongoing change in Rioplatense Spanish (Argentina and Uruguay), from *yeismo*, to *zheismo*, to *sheismo*. The historical palatal lateral of Castilian has merged with the palatal glide in most of Latin America (*yeismo*), but in Rioplatense this glide changed to a voiced palatal fricative by the 20th Century (*zheismo*). Now, it is devoicing to /ʃ/, so that a phrase like 'yo me llamo' is /ʃo me ʃamo/ for the average young Porteña or Uruguayan (/uruguaiʃa/). This change has been the subject of several well-known studies, by Fontanella de Weinberg (1979), Donni di Mirande (1992), and others. Table 1 shows results from Rohena Madrazo's (2007) study of Buenos Aires, focusing on the last step in this chain, i.e., rates of devoicing. His younger speakers devoice at about twice the rate of older speakers. These results also show the classic social distribution of a spontaneous change in progress: women lead over men, and the change is most advanced in the middle and working classes, with the upper class lagging.

Table 1. Devoicing of palatal fricatives in Buenos Aires Spanish.
(Factor weights, percentages and Ns; from Rohena Madrazo 2007)

Gender	Females	.59	54%	407/751
	Males	.41	48%	347/709
Age	Younger Speakers	.66	66%	492/741
	Older Speakers	.34	36%	262/719
Social Class Index	I	.19	20%	110/543
	II	.74	71%	455/639
	III	.62	67%	189/278
Totals		**Input .50**	**51%**	**754/1460**

Brazilian Portuguese is also undergoing a number of important spontaneous changes. A remarkable development is the emergence of a new first person plural pronoun. The historical *nós*, dating from before Latin, has over about the last 50 years been almost completely replaced by *a gente*. This began as a noun phrase meaning 'the people', but now is so entrenched in the meaning 'we' that it is almost ungrammatical as a noun.

Table 2 shows Zilles' (2005) findings on speakers from Porto Alegre, drawn from the Varsul corpus. Her speakers over the age of 50 already use *a gente* 65% of the time, but her younger age group, 25–50 year olds, are at 78%. The other classic signs of a change in progress are also there: women lead men, 72% to 62%, and the social class distribution, here represented by educational levels, peaks in the middle, with 74% usage among those with a high school education, but less at both extremes, those with primary school or college. More recent data show the change is still advancing: in the São Paulo 2010 corpus (GESOL-USP n.d.) collected some two decades after the Varsul corpus, I have found young speakers with 85% usage of *a gente*. At the current rate of change, *a gente* may completely replace *nós* in the spoken language within a few generations.

Table 2. Grammaticalization of a gente 'we' in Brazilian Portuguese.
(Factor weights, percentages and Ns. Zilles 2005)

Gender	Females	.55	72%	915/1266
	Males	.41	62%	422/678
Age	25–50	.66	78%	480/618
	50–70	.42	65%	857/1326
Level of education	Elementary	.33	54%	128/238
	Intermediate	.55	74%	411/559
	Secondary	.50	74%	315/425
	Post-secondary	.52	67%	483/722
Totals		**Input .85**	**69%**	**1337/1944**

3. Parallel changes in Caribbean Spanish and Brazilian Portuguese

A remarkable development in New World Romance is the suite of shared innovations that are found in the popular vernaculars of Caribbean Spanish (CS) and Brazilian Portuguese (BP). Each of these varieties is substantially different from other dialects of the same language, but what is striking is that the changes that both have undergone are very similar. Thus we have Western Hemisphere dialects of different languages from different European sources that are **not** in contact with each other, but share an extensive collection of linguistic innovations. This coincidence demands an explanation. Let us consider some of these shared features.

3.1 Phonological reductions

Some of the shared features are phonological. Both CS and BP have extensive patterns of coda consonant reductions. All the possible syllable-final consonants, especially word-final consonants, occur with lenition or complete deletion in both languages. In Spanish, this includes all coronal consonants:

(1) a. /-s/ *estamos, las casas*
b. /-d/ *ciudad, usted*
c. /-n/ *bien, hablan, nación*
d. /-l/ *mal, papel*
e. /-r/ *mar, sentir, mujer*

In Brazilian Portuguese the inventory is smaller, consisting mainly of /-s/ (variably realized as /s, z, ʃ, ʒ/ depending on dialect and context) and /-r/ (also with great dialectal variation):

(2) a. /-s/ *estamos, as casas*
b. /-r/ *mar, sentir, mulher*

In popular BP and CS, all of these coda consonants can be lenited or deleted. The coda /-s/ case is the most-studied, and functions as a sociolinguistic marker: it is popularly described in the Caribbean as 'eating the esses' (*comerse los esse*'). When /-s/ is not deleted in CS, it is often realized as a lenited or aspirated form, with a lax posterior lingual or glottal constriction [x, h]. BP mostly shows deletion, but aspiration is found, especially before nasals or /l/ (*mehmo*).

Numerous quantitative studies of popular varieties of CS and BP demonstrate high rates of /-s/ lenition and deletion. In Panama Cedergren (1973) found 59% deletion overall, rising to 68% in the lowest socioeconomic class; Alba's (1990) study of Dominican Spanish reported 71% deletion among male speakers and 51% deletion among females. In Brazil, Guy (1981) found a deletion rate of 53% of unstressed non-inflectional -s among illiterate speakers in Rio de Janeiro (RJ).

All such studies report substantial differences in deletion rates and constraints between roots and suffixes (*menos* vs. *casas*). This is treated below in terms of morphosyntactic variation.

Table 3. Coda consonant deletion in Caribbean Spanish and Brazilian Portuguese

	-s deletion
Panama (Cedergren 1973)	59%
Dominican Republic (males, Alba 1990)	71%
Brazil-RJ (*analfabetos*, Guy 1981)	53%
	-r deletion
Dominican Republic (LC, Alba 1990)	68%
Brazil-SP (Oushiro & Mendes 2013)	55%

Coda /r/ is similarly deleted at high rates in popular dialects of CS and BP. Oushiro and Mendes (2013) report 55% -r deletion in the SP 2010 corpus, and Alba reports 68% deletion among low-income speakers in the DR. In both languages -r deletion is more common in verbal infinitives, where -r is an affix (*estar, comer, sentir*). Alba finds -r deleted in 48% of infinitives vs. 32% of other words in his full corpus.

3.2 Morphosyntactic changes

The popular varieties of CS and BP are systematically different from their standard counterparts and historical European sources, but similar to each other, in a number of features of the morphosyntax. Some important examples are: subject pronoun expression, nominal and verbal number agreement, and repeated negation.

3.2.1 *Subject pronoun expression*

Portuguese and Spanish, like Latin, are null subject languages; i.e. a sentence without an overt subject is perfectly grammatical. For example, one of the São Paulo 2010 speakers says within the space of one minute both *"Eu não sou contra"* and *"Não sou contra"* first with and then without the subject pronoun *eu*. In European varieties of Spanish and Portuguese expressed pronominal subjects are fairly rare, with quantitative studies finding rates such as 25% in Madrid Spanish (Cameron 1993), and 22% in European Portuguese (Barbosa et al. 2005). In non-Caribbean dialects in Hispanic America, similar figures are found, but in the Caribbean region, these rates are considerably higher.

Illustrative figures are given in Table 4 from Otheguy and Zentella's (2012) definitive study of Spanish speakers in New York City. Their 'mainlanders' from Mexico, Ecuador and Colombia, use 28% expressed pronominal subjects (the

Mexicans use only 22%), while their Caribbean speakers average 39%, with the Dominicans peaking at 41%. Otheguy and Zentella further report that longer stays in NYC are associated with higher pronoun rates, possibly as a consequence of contact with English. But the substantial difference between *Caribeños* and *continentales* is present regardless of length of stay.

Table 4. Subject pronoun expression by New York City Spanish speakers (Otheguy & Zentella 2012)

Place of origin	speakers	% overt subject pronouns
Caribbean (Cuba, Puerto Rico, Dominican Republic)	72	39
Mainland (Mexico, Colombia, Ecuador)	68	28

Another perspective is provided by Rafael Orozco (2018) (see Table 5), who compared subject pronoun expression among speakers from Barranquilla, on the Caribbean coast of Colombia, with *barranquilleros* living in New York City. Both are significantly higher than speakers from Spain or Mexico or Ecuador, or from the Colombian interior. In fact, the two groups bracket Otheguy and Zentella's figures for other Caribbean speakers in New York. Orozco's results can be taken as a rough indicator of the effect of the language contact his subjects encounter in New York: they show increased use of overt pronouns, but start from a higher base rate than the non-Caribbean dialects.

Table 5. Subject pronoun expression by Spanish speakers in Barranquilla and *barranquilleros* in New York (Orozco 2018)

Place of residence	N	% overt subject pronouns
Barranquilla	1031/3009	34.3
New York City	1303/3006	43.3
		$p = .000$

In Brazil, subject pronoun expression is strikingly higher than even the Hispanic Caribbean. Table 6 presents a sample of representative results, from Duarte (1993, 1995) and Gilbert, Guy and Robinson (2017).

Duarte (1993) conducted a historical study of subject pronouns in Brazilian popular drama (see Figure 1). The rate of overt subjects swells from a low European level in the earliest period, to 74% in the most recent data. Combining these results with those presented above, the overall picture is that CS and BP are two to three times more likely to use an overt subject pronoun than their European counterparts, and in the case of CS, appreciably more likely than other American varieties of Spanish.

Table 6. Subject pronoun expression in Brazilian and European Portuguese

Variety/source	% overt subject pronouns
Portugal – Duarte 1995	34
Brazil – Duarte 1993	75
Brazil – Duarte 1995	71
Brazil – Gilbert, Guy & Robinson 2017	65

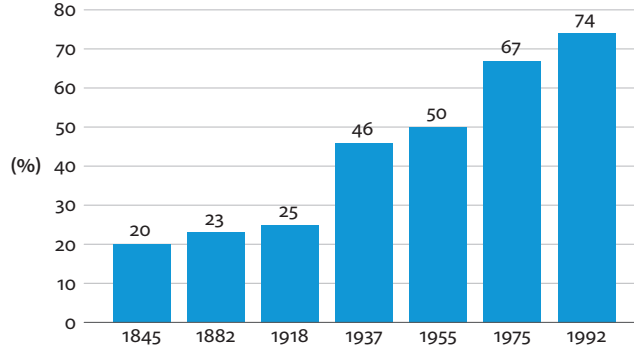


Figure 1. Overt pronominal subjects through seven periods

3.2.2 *Number agreement*

Variable or absent number agreement is another significant innovation in CS and BP which differentiates them from other dialects of Spanish and Portuguese. In noun phrases, Spanish and Portuguese have had obligatory number agreement for determiners, nouns and adjectives, like their source, Latin (and Proto-Indo-European). But in contemporary BP and CS, number marking is variable, and many historically obligatory number markers are now omitted. Notably, the syntactic position of where markers occur is strongly constrained in both languages. The first word is almost always marked, but subsequent words are progressively less likely to bear a plural marker. Since the first word is usually a determiner, these results are consistent with the argument advanced by Cyrino and Espinal in this volume, that Number is specified on the determiner, not on the noun.

Many studies have looked at this variable in CS and BP. Figures from Terrell's (1979) study of Cuban Spanish are given in Table 7. First position modifiers are almost categorical at 97% marking, while non-initial nouns and adjectives are at 60%. Similar conditioning is found in BP: Guy 1981 reports 95% overt markers in first position, dropping to 21–28% in later positions (see Table 8).

Table 7. Nominal plural marking by syntactic position in Cuban Spanish (Terrell 1979)

Position/word class	% plural marked
First – modifiers	97
First – nouns	81
Non-initial – nouns	61
Post-nominal – adjectives	60

Table 8. Plural marking by syntactic position in RJ Portuguese (Guy 1981)

Position in NP:	N	% plural-marked
first	5247	95
second	3947	28
third	552	21

This constraint on plural marking, favoring maintenance at the beginning of NPs and absence towards the end, is a novel development in the Romance languages. It has been confirmed by many studies for both Caribbean Spanish and Brazilian Portuguese, but it is completely absent in European Spanish and Portuguese, and in Mexican and Andean Spanish. So the obvious challenge for an explanatory linguistics is: why do BP and CS share this constraint, while all their other linguistic relatives lack it?

The same question arises with variable verbal number agreement. In both BP and CS, 3rd person plural subjects can be found with 3rd singular verbs. This phenomenon has been most studied in BP. The most striking finding of this research is what Naro and Lemle (1976), Naro 1981 dubbed the 'saliency constraint'. That is, the verb forms with the most distinctive differences between singular and plural are the ones that are most likely to be plural-marked.

The relevant figures from Guy 1981 appear in Table 9. The least salient cases are those where singular and plural are distinguished only by an oral vs. nasal vowel: *come-comem*. As we move up the salience scale, the differences become more marked, culminating in irregular preterites and suppletive alternants, such as *é-são* and *fez-fizeram*. As Table 9 shows, the rate of plural marking is three to five times greater in the most salient categories than the least salient.

In Spanish, this phenomenon is less well documented, but Poplack 1980 reports more plural marking in preterites and irregular verbs than in simple present tense forms like *hablan*. So again there is a conundrum: what change process led to more plural marking where it is more noticeable, and fewer where it is less noticeable, and why did this occur in the Caribbean and Brazil, but not elsewhere in the Spanish- and Portuguese-speaking regions of the world?

Table 9. Verbal plural marking in RJ Portuguese, by saliency (Guy 1981)

Category	N	% Plural marked
1. *comem-comem, fale-falem*, etc.	894	14
2. *fala-falam, ia-iam*, etc.	3161	26
3. *faz-fazem, quer-querem*, etc.	481	30
4. *dá-dão, está-estão*, etc.	1112	63
5. *sumiu-sumiram, foi-foram*, etc.	476	66
6. *falou-falaram, fez-fizeram, é-são*, etc.	1776	76

3.2.3 *Negative repetition*

Another shared morphosyntactic innovation in CS and BP is reduplicated or repeated negation, a distinctive pattern in which the usual preverbal negator is repeated again at the end of the clause. Bullock and Toribio (2009) provide Dominican examples:

(3) a. Para acá **no** ha llovido nada **no**.
b. Los haitianos **no** hablan como nosotros **no**.

Some Portuguese examples from Cavalcante 2009 are:

(4) a. **Não** gosta de ir em festa **não**.
b. Eu **não** sei do quê que ele tá trabalhando lá **não**.

In Brazil this reduplicated negation is widespread and common (22% of all negative constructions in Reimann & Yacovenco 2011, 30% in Cavalcante 2009). In Hispanic America it is found in the Caribbean and sporadically elsewhere. Parallel structures are notably evident in Spanish- and Portuguese-based creoles, such as Palenquero and São Tomense (Schwegler 1991; Ferraz 1979; Hagemeijer 2008), and in *bozal* Spanish in Cuba and elsewhere. For Palenquero and São Tomense, Schwegler and Ferraz argue for a West African substrate as the source of this construction, as it is found throughout a number of languages from Southwest Africa to the Congo, including those spoken by many of the people taken to the Americas by the slave trade.

4. Explanation

An adequate account of the history of Spanish and Portuguese in the Americas should explain the contemporary sociolinguistic facts in Brazil and the Caribbean sketched above. There are a number of related questions: (a) Why do we find these specific features? (b) Why do we find them in CS and Brazilian Portuguese, and

not elsewhere? (c) Where did they come from? d) Why did they appear at one historical time, not before or after?

The explanations that are proposed for these facts fall into three clusters attributing them to the following sources: (1) European roots; (2) independent spontaneous innovation; (3) language contact and possibly creolization. Of these, I dismiss independent innovation at the outset. It is wildly improbable that these two varieties, not in contact with each other, each just happened to come up with the same unusual innovations at the same time in history. So I focus on the other two possible explanations.

4.1 European sources for popular American varieties

First, did they come from Europe? Much linguistic work seeks to explain linguistic traits of colonial or frontier communities in terms of their homeland sources (cf. Trudgill 2004). Among the features considered here, Andalusian sources have been proposed for -s lenition in CS, but the parallel phenomena in BP mostly lack a peninsular source. Peninsular explanations for the phonology of CS and BP are also inconsistent with the dialect geography. There was no population movement from Iberia to America that explains why these phenomena are found where they are, and not elsewhere, and mainly occur among poor black speakers rather than affluent whites.

As for the unusual morphosyntactic traits of BP and CS, peninsular sources are virtually nonexistent. Negative repetition is rare in Iberia and robust absence of plural markers occurs nowhere in Spain or Portugal. The constraint effects seen above – initial word marking in nominal plurals, saliency constraints in verbal plurals – are likewise unattested in Spain. The only serious proposal of a European source for such phenomena is the work of Naro and Scherre (1993, 2000) on peninsular Portuguese. These scholars report having observed non-agreeing forms in Portugal, but they concede that such forms are "statistically rare" (2000: 241) and that the Portuguese linguists they consulted all denied such forms existed. Examining texts written before the 17thC, they found a few cases of singular verbs with plural subjects, amounting to less than 1% of all relevant cases, among which they claimed to find a saliency constraint. On this basis they argue that agreement loss was incipient in preclassical Portuguese, and happened to drastically expand in Brazil because of the "conditions of endemic pidginization and adult second language acquisition." This claim has been contradicted in the literature by Vieira and Bazenga (to appear), who find that their data "do not sustain the proposition that there was, in the development of Brazilian Portuguese, the expansion of European structures." Hence I will argue that the BP pattern arises directly from the processes of 'pidginization and adult second language acquisition' that Naro and Scherre allude to, rather than from a non-existent peninsular model.

In any case, the fatal flaw for any account that seeks European sources for the phenomena considered here is that no single European source explains the extensive parallels between CS and BP. For most of their history, the Spanish and Portuguese empires were separate and frequently hostile operations, with little cross-settlement or linguistic contact. Hence if separate settler streams from Spain and Portugal implanted separate languages in separate territories, any idiosyncrasies shared between American varieties of the languages must have a non-European source. We turn, then to the matter of language contact.

4.2 Language contact

Many linguistic phenomena occur in language contact; here our concern is phenomena associated with adult second language acquisition and cross-language transference. Spanish and Portuguese in the Americas experienced language contact on a massive scale – first with indigenous languages, and subsequently with the languages of millions of Africans. A majority of the ancestors of the people of contemporary Latin America did not come from Spain and Portugal, and most of them must have learned Spanish or Portuguese, or a contact-influenced variety thereof, as adults. Since adults are notoriously poor second language learners, the varieties of Spanish/Portuguese acquired by these learners certainly differed from the 'target' varieties spoken by European settlers. Some of these differences reflect transfers or impositions from the speaker's L1. Other differences reflect communicative strategies of early-stage adult L2 learners, including generalizations, focusing on roots to the exclusion of inflections, avoidance of movement phenomena, etc. If language contact and 'irregular transmission' (Lucchesi 2003, cf. also Naro & Scherre 2003) are the source of CS/BP features, they should reflect L1 impositions and L2 acquisition strategies.

The most divergent linguistic effect of language contact is the emergence of pidgin/creole varieties. These are associated with extreme social conditions, namely those found (a) in enslaved communities composed of people with diverse language backgrounds, who (b) undergo abrupt, coerced language shift, and (c) have limited contact with L1 speakers of the dominant target language, because (d) the enslaved speakers greatly outnumber the TL speakers, and (e) are socially segregated.

These processes are what gave rise to Luso-Hispanic creole languages such as Palenquero and Papiamentu, and to the many English- and French- lexifier creoles in the Caribbean. Given the massive presence of enslaved Africans in the founding populations of CS and BP, we should ask whether these varieties started out as creoles. There are a number of Portuguese based creoles in Africa such as Kaboverdianu and the several creoles of São Tomé. In light of the Palenquero case – a Spanish-based creole spoken by the descendants of an ex-slave community in Colombia – creoles must have been spoken in the many such communities,

palenques and *quilombos*, that existed throughout Latin America. But linguistic divergence is a matter of degree, and there are less extreme linguistic outcomes in contact situations that involve less extreme cases of the social conditions cited above as drivers of creolization, such as a higher proportion of Europeans in the colonial population, or less extreme social segregation between blacks and whites. The popular varieties of CS or BP don't have to be labeled as creoles to recognize that the enslavement of a substantial fraction of their founders was unlikely to promote acquisition of a highly convergent variety. The varieties that emerged among enslaved peoples were divergent by virtue of language contact and irregular transmission, and across generations they continued to differ from the European norm to the extent that people of African descent continued to be socially and linguistically separated from the white elite. This is the likely sociolinguistic history that gave rise to popular varieties of Spanish and Portuguese spoken mainly by people of African descent in Latin America. Therefore, we should inquire whether the distinctive shared properties of CS and BP that we have identified have their sources in African influences or adult L2 learner strategies.

4.3 African sources

Several of the phenomena observed in CS and BP have likely African sources. The vast majority of people taken by the slave trade to the Americas spoke languages of the Niger-Congo family. A systematic phonological trait of these languages is the preponderance of open syllables. Proto-Bantu, and probably proto-Niger-Congo, permitted only CV or V syllables, and this is still true of most languages in this family. Coda consonants are nearly or entirely absent in most of the languages spoken by the Africans transported to the Caribbean and Brazil (although some permit nasal elements in the syllable nucleus or coda: syllabic nasals, nasal vowels, occasionally coda nasal consonants). This is true of Yoruba, Igbo, Ewe, Kimbundu, Umbundu, Kikongo, etc. Hence the 'foreign accent' imposed by African speakers on the varieties of Spanish or Portuguese they acquired in the Americas would surely have omitted or lenited coda consonants. Insofar as this served as the model for their children and subsequent generations, it is likely to have been a significantly greater contributor to coda lenition in Luso-Hispanic America than Andalusian Spanish or other European sources. Crucially, the phonological similarity of the African source languages is the only input factor common to the founding populations of both CS and BP. Reduction of all coda consonants would be expected in a population undergoing shift from Niger-Congo languages to Spanish and Portuguese, but would be a statistically improbable coincidence were it a novel innovation arising spontaneously and simultaneously in two different languages in the Americas.

Negative repetition also has likely African sources; as we have noted it is found not only in BP and CS, but also in Luso-Hispanic creoles (e.g. São Tomé creoles and Palenquero), and is attributed by specialists in these languages to a West African substrate (Ferraz 1979; Schwegler 1991). Schwegler specifically cites KiKongo where negation involves two negative particles, one preverbal and the other clause-final.

Marking nominal plurality in phrase-initial or pre-head position is an unusual innovation from an Ibero-Romance perspective, but has striking parallels with many West African languages. The most prominent African linguistic groups represented in Brazil were drawn from two areas: the Nigerian coast (especially Yoruba and Igbo), and the area from the Congo basin through present-day Angola (especially the Bantu languages Kimbundu, Umbundu, and Kikongo). Portuguese and Brazilian slavers had well-developed operations and local connections in these regions, and were also heavily involved in supplying slaves from these areas to the Hispanic Caribbean. Most of these languages share the very structural property in their nominal plural marking systems that we find in CS and BP: plural markers are located in phrase-initial or pre-head position. In Bantu languages, including Kimbundu and Kikongo, plural marking is accomplished by a prefix; hence any plural NP has an initial morpheme indicating plurality. Among the non-Bantu west African languages, we find cases like Yoruba, which marks plurals with an NP initial plural determiner, *awon*, homophonous with the 3PL pronoun.

Given these structural facts, the likely source of the position constraint on NP plural marking is transference from the African languages spoken by significant majorities in colonial times. Adult L2 learners of Spanish and Portuguese, whose L1s had NP-initial plural marking, imposed this structure on their L2. Their L1 syntax primed them to perceive and produce plural markers at the beginnings of NPs, before the head noun. Analogous transference effects can be noted among English speakers learning Spanish or Portuguese: they impose English structures, placing plural markers on nouns but not adjectives, locating adjectives pre-nominally, etc.

4.4 Adult learner strategies

Other features considered above are consistent with more general strategies of adult L2 learners. When obliged to use the L2 for real-life communicative ends, early stage learners tend to utilize: (a) frequent structures more than rare ones, (b) salient forms more than unsalient (e.g., stressed vs. unstressed forms), (c) roots to the detriment of inflected forms, and (d) free morphemes over bound ones. Much of what makes BP and CS distinctive is consistent with such strategies. The saliency constraints suggest acquisition of agreement marking by untutored

learners in verbal forms where it was most audibly prominent. High rates of omission of inflectional -r, -s, -n, found consistently in BP and CS but absent in Spain and Portugal, are consistent with the focus on roots. Inflectional loss is reflected in other ways: the BP verbal paradigm is often described as highly impoverished: along with variably absent plural marking in third person forms, the second person forms are almost entirely absent in most of the country, and the 1PL form is frequently replaced by a construction using a 3SG verb with *a gente* as the subject. Consequently, it is unremarkable to encounter speakers who produce for example the imperfect paradigm with no person-number inflection at all: *eu, você, ele, a gente, vocês, eles falava*. Loss of other low-frequency structures is also evident, such as the complete loss of mesoclitics in BP.

The high rate of subject pronoun expression in BP and CS is also consistent with irregular transmission. Creoles systematically prefer free morphemes over bound or null morphemes, they lack person number inflection on verbs, and almost always require obligatory overt subject pronouns, much like English. Hence we may conjecture that the original African speakers in the Americas would have had high rates of overt subject pronouns in the varieties of Spanish and Portuguese that they were acquiring or creating, alongside minimal or no person/number marking on verbs. The fact that rates of subject pronoun expression are now high but not categorical reflects a history of irregular transmission followed by influence from continuing contact with the standard language, which still allows null subjects.

What, then, about Duarte's evidence of increasing use of overt SPs in the 19th and 20th centuries? I argue that this does not necessarily reflect the progress of a local and ongoing innovation in the language, but rather a change in literary norms. Recall that her data are drawn from the texts of plays. 19th C literary traditions in Brazil emphasized the usage of elite speakers and emulated European Portuguese norms. The 20th century saw an increasingly Brazilianized literary language, converging today on something closer to vernacular usage, like the São Paulo speakers studied by Gilbert et al. 2017.

4.5 Afro-Bolivian Spanish

Another crucial piece of evidence that supports the theory that the CS/BP features derive from African language contact is the case of Afro-Bolivian Spanish, recently described by several scholars, including Lipski (2008). Few African slaves were taken to the Andean highlands, but one group ended up in Bolivia in the Yungas on the eastern slopes of Andes, where they lived in a state of bondage or peonage until the mid-20th century. Their overseers were mostly Aymara-speaking indigenous peoples, and the peons were typically obliged to speak Aymara with the

overseers. Consequently these Afro-Bolivians had little continuing contact with Spanish for several hundred years until the second half of the 20th century. Hence their language was not subject to much standardization or decreolization and provides a close perspective on the speech of *bozales* and other early generations of Africans in Spanish America. The linguistic characteristics of BP and CS are clearly evident in this Afro-Bolivian community, often in more pronounced form. These include:

- *coda consonant reduction.* According to Lipski, ABS has extensive deletion or aspiration of final -s, and deletion of final -r, especially in verbal infinitives. These characteristics are in striking contrast to the consonantally conservative highland Bolivian Spanish.
- *nominal plural marking.* Nouns and adjectives are mostly invariant in this variety, bearing no plural markers. In the most traditional usage, plural NPs are marked by an initial plural determiner *lu* or an initial numeral: e.g. *lu mujé* 'las mujeres', *tres hombre* 'tres hombres'. But in what Lipski (2008: 93) describes as "partially modernized speech" the plural structure consists of "marking plural -s only on the *first* element of plural noun phrases".
- *verbal plural marking.* Lipski highlights the ABS verbal system as one of its most distinctively 'restructured' characteristics. It shows no person-number inflection at all. The invariant verb form occurring with all subject persons and numbers is based on the standard Spanish 3SG, as in Table 10.

Table 10. Invariant person/number marking in Afro-Bolivian

Afro-Bolivian	Spanish	Gloss
yo tiene	yo tengo	'I have'
nojotro tiene	nosotros tenemos	'we have'
ele tiene	el tiene	'he has'
eyu tiene	ellos tienen	'they have'

In the 'partially modernized' variety of ABS Lipski observes that "the most frequent verbs (usually irregular) are the first to exhibit subject-verb agreement" (p. 110). He cites examples such as *yo sé, soy, tengo* 'I know, I am, I have' instead of *sabe, es, tiene*. All of these cases would count as highly salient in the scales used in the analysis of BP.

- *Subject pronouns*. Finally, ABS has nearly categorical use of overt subject pronouns. This places it, like BP, very close to non-null subject languages like modern English, and a very long way from most dialects of Spanish outside

of the Caribbean. By comparison, Highland Bolivian Spanish is like Mexican and peninsular Spanish in using very low rates of subject pronouns.

In all these respects, the language of Afro-Bolivians is strikingly different from the highland Bolivian dialect of Spanish that is spoken around it. Afro-Bolivians have been isolated for centuries from the Spanish varieties with which they share the features considered here. There is no plausible scenario in which they could have acquired them from European speakers or spontaneously developed them. Rather, these features must have emerged from the conditions of enslavement, language contact, and language acquisition. Lipski concludes that ABS is descended from a proto-Hispanic pidgin/creole, spoken originally in the Caribbean. The parallels with BP and CS certainly suggest common processes at work. The greater distinctiveness of ABS locates it farther from the superstrate language, and closer to the constellation of characteristics we find in generally recognized creoles. Its history of linguistic isolation implies that ABS must be more basilectal, closer to the speech of the earliest generations of Africans in the Americas, than BP and CS. This in turn implies a historical trajectory by which all of these varieties started out as creoles, or at least divergent restructured varieties, and then acquired their present form through differing degrees of standardizing influences.

In sum, the distinctive linguistic characteristics shared by BP and CS and Afro-Bolivian Spanish, that set them apart from their peninsular source languages, all have consistent explanations in terms of African substrates and/or adult second language acquisition. They all originated in communities where a substantial African population outnumbered the European populations by a ratio of two or three to one. Given that some six million Africans were transported to Luso-Hispanic America and obliged to communicate in Spanish or Portuguese, such linguistic developments are not the least bit surprising. It is more surprising that there exists such a long-standing and ample scholarly tradition which ignores or minimizes such effects.

5. Conclusion

The American varieties of Spanish and Portuguese arrived at their contemporary forms following several different trajectories, driven by different sociolinguistic processes. Much of what they are today came from Europe, but in the New World, they were thrust into a simmering linguistic stew, from which they emerged transformed. Some of the major transformations arose from contact with many other languages, whose speakers found it necessary to shift to Spanish and Portuguese under a range of conditions, most of them adverse, and thereby transforming the

varieties that they acquired. Countervailing conservative forces have exerted pressures to reduce divergence. But ongoing change persists, producing novel innovations that differ from place to place. As for the future trajectories of Western Hemisphere Spanish and Portuguese, the one thing we can be sure of is that they will continue to change.

References

Alba, Orlando. 1990. *Variación fonética y diversidad social en el español dominicano de Santiago*. Santiago, DR: Universidad Católica Madre y Maestra

Barbosa, Pilar, Maria Eugênia Duarte, and 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

Baxter, Alan. 1997. "Creole-Like Features in the Verb System of an Afro-Brazilian Variety of Portuguese." In *The Structure and Status of Pidgins and Creoles*, ed. by Spears, Arthur, and Donald Winford, 265–88. Amsterdam & Philadelphia: Benjamins. https://doi.org/10.1075/cll.19.16bax

Baxter, Alan. 2009. "A concordância de número." In Lucchesi et al., 269–294.

Bullock, Barbara E., and Almeida Jacqueline Toribio, 2009. "Reconsidering Dominican Spanish: Data from the Rural Cibao." *Revista Internacional de Lingüística Iberoamericana* 7 (2): 49–73.

Callou, Dinah, João Moraes, and Yonne Leite. 1996. "Variação e diferenciação dialetal: a pronúncia do /r/ no português do Brasil." In *Gramática do português falado, Vol. VI*. Campinas: Editora da Unicamp.

Cameron, Richard. 1993. "Ambiguous Agreement, Functional Compensation, and Nonspecific tú in the Spanish of San Juan, Puerto Rico and Madrid, Spain." *Language Variation and Change* 5: 305–334. https://doi.org/10.1017/S0954394500001526

Cavalcante, Rerisson. 2009. "A negação sentencial." In Lucchesi et al., 251–268.

Cedergren, Henrietta. 1973. "The Interplay of Social and Linguistic Factors in Panama". Doctoral dissertation, Cornell University.

Clements, J. Clancy. 2009. *The Linguistic Legacy of Spanish and Portuguese: Colonial Expansion and Language Change*. Cambridge: Cambridge University Press.

Cyrino, Sonia, and M. Teresa Espinal. in this volume. "Number as an Adjunct in Romance".

De Oliveira, Marco Antônio. 1983. "Phonological Variation and Change in Brazilian Portuguese: The Case of the Liquids". Doctoral dissertation, University of Pennsylvania. http://repository.upenn.edu/dissertations/AAI8406656

Donni de Mirande, Nélida Esther. 1992. "Sobre el ensordecimiento del žeísmo en Rosario (Argentina)." In *Homenaje a Humberto López Morales*, ed. by María Vaquero, and Amparo Morales, 171–183. Madrid: Arco Libros

Duarte, M. E. 1993. Do pronome nulo ao pronome pleno: A trajetória do sujeito no português do Brasil." In *Português Brasileiro: uma viagem diacrônica. Homenagem a Fernando Tarallo*, 107–128. Campinas, Brazil: Editora da UNICAMP.

Duarte, M. E. 1995. A perda do princípio "Evite Pronome" no português brasileiro." PhD Dissertation, Campinas, Brazil: UNICAMP .

Duarte, M. E. 2003. A evolução na representação do sujeito pronominal em dois tempos." In *Mudança linguística em tempo real*, ed. by Paiva, and Duarte. Rio de Janeiro: Contra Capa.
Fontanella de Weinberg, María B. 1979. "Un cambio lingüístico en marcha: las palatales en el español bonaerense". *Orbis*: 27: 215–47
Ferraz, Luis. 1979. *The Creole of São Tomé*. Johannesburg: Witwatersrand University Press.
Friedemann, Nina S. de, and Carlos Patiño Roselli. 1983. *Lengua y sociedad en El Palenque de San Basílio*. Bogotá: Instituto Caro y Cuervo.
GESOL-Grupo de Estudos e Pesquisa em Sociolinguística. n.d. *Projeto SP2010*. Universidade de São Paulo. http://projetosp2010.fflch.usp.br/node/1
Gilbert, Madeleine, Gregory R. Guy, and Mary Robinson. 2017. *Subject Pronoun Expression in Brazilian Portuguese*. Presentation at NWAV 46 – Annual Conference on New Ways of Analyzing Variation, University of Wisconsin.
de Granda Gutiérrez, Germán. 1978. *Estudios lingüísticos hispánicos, afrohispánicos y criollos*. Madrid: Gredos.
Greenberg, Joseph. 1987. *Language in the Americas*. Stanford: Stanford University Press.
Guy, Gregory. 1981. "Linguistic Variation in Brazilian Portuguese: Aspects of the Phonology, Syntax, and Language History". Doctoral dissertation, University of Pennsylvania. http://repository.upenn.edu/dissertations/AAI8117786
Guy, Gregory. 1990. "The Sociolinguistic Types of Language Change." *Diachronica* 7: 47–67. https://doi.org/10.1075/dia.7.1.04guy
Hagemeijer, Tjerk. 2008. "Double-Headed Negation in Santome." *Journal of Portuguese Linguistics* 7 (2): 63–82. https://doi.org/10.5334/jpl.128
Hoffman, Michol. 2010. *Salvadorian Spanish in Toronto* (Lincom Studies in Romance Linguistics 63). Munich: Lincom Gmbh.
Holm, John. 1992. "Popular Brazilian Portuguese: A Semi-Creole." In *Actas do colóquio sobre crioulos de base lexical portuguesa*, ed. by Ernesto d'Andrade, and Alain Kihm, 37–66. Lisboa: Colibri.
Holm, John. 2000. "Semi-Creolization: Problems in the Development of Theory." In *Degrees of Restructuring in Creole Languages*, ed. by I. Neumann-Holzschuh, and E. Schneider, 19–40.
Jacobs, Bart. 2012. *Origins of a Creole: The History of Papiamentu and its African Ties*. Berlin: De Gruyter. https://doi.org/10.1515/9781614511076
Labov, William. 1972. "The Social Stratification of (r) in New York City Department Stores." In *Sociolinguistic Patterns*, ed. by W. Labov, 43–54. Philadelphia: University of Pennsylvania Press.
Lipski, John. 2008. *Afro-Bolivian Spanish*. Madrid: Iberoamericana, and Frankfurt: Vervuert. https://doi.org/10.31819/9783865279026
Lovejoy, Paul E. 1994. The Volume of the Atlantic Slave Trade. A Synthesis." In *The Atlantic Slave Trade*, ed. by David Northrup. D.C. Heath and Company.
Lovejoy, Paul E. 2000. *Transformations in Slavery*. Cambridge University Press.
Lucchesi, Dante. 2003. "O conceito de transmissão irregular e o processo de formação do português do Brasil." In Roncarati, and Abraçado, 272–84.
Lucchesi, Dante, Alan Baxter, and Ilza Ribeiro, 2009. *O Português Afro-brasileiro*. Salvador: EDUFBA (Editora da Universidade Federal da Bahia). https://doi.org/10.7476/9788523208752
McWhorter, John. 2000. *The Missing Spanish Creoles: Recovering the Birth of Plantation Contact Languages*. Berkeley: University of California Press.

Morton, Thomas. 2005. "Sociolinguistic Variation and Language Change in El Palenque de San Basílio, Colombia". Doctoral dissertation, University of Pennsylvania. http:// repository.upenn.edu/dissertations/AAI3197717

Naro, Anthony J. 1981. "The Social and Structural Dimensions of a Syntactic Change." *Language* 57 (1): 63–98. https://doi.org/10.1353/lan.1981.0020

Naro, Anthony J., and Miriam Lemle. 1976. "Syntactic Diffusion." In *Papers from the Parasession on Diachronic Syntax*, ed. by S. Steever, C. Walker, and S. Mufwene, 221–40. Chicago: CLS.

Naro, Anthony J., and Maria Marta P. Scherre 1993. "Sobre as origens do português popular do Brasil." *Revista DELTA*, 9: 437–454.

Naro, Anthony J., and Maria Marta P. Scherre. 2000. "Variable Concord in Portuguese." In *Language Change and Language Contact in Pidgins and Creoles*, ed. by J. McWhorter. Amsterdam and Philadelphia: John Benjamins.

Naro, Anthony J., and Maria Marta P. Scherre. 2003. "O conceito de transmissão irregular e as origens estruturais do português brasileiro: um tema em debate." In Roncarati, and Abraçado, 285–302.

Orozco, Rafael. 2018. *Spanish in Colombia and New York City: Language Contact Meets Dialectal Convergence*. Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/impact.46

Otheguy, Ricardo, and Ana Celia Zentella. 2012. *Spanish in New York: Language Contact, Dialectal Leveling, and Structural Continuity*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199737406.001.0001

Oushiro, Livia, and Ronald Mendes. 2013. "A pronúncia do (-R) em coda silábica no português paulistano." *São Paulo: Revista do GEL* 8 (2): 66–95.

Poplack, Shana. 1979. "Function and Process in a Variable Phonology". Doctoral dissertation, University of Pennsylvania.

Poplack, Shana. 1980. "On Deletion and Disambiguation in Puerto Rican Spanish." *Language* 56 (2): 371–385. https://doi.org/10.1353/lan.1980.0033

Reimann, Cristiana A., and Lilian Yacovenco. 2011. "A dupla negação no português falado em Vitória/ES: traço da identidade linguística capixaba? *Anais do congresso nacional de estudos lingüísticos*. Vitória: CONEL.

Richardson, Bonham C. 1992. *The Caribbean in the Wider World, 1492–1992. A Regional Geography*. Cambridge University Press. p. 166. https://doi.org/10.1017/CBO9780511560057

Rohena Madrazo, Marcos. 2007. *Devoicing of Palatal Fricatives in Buenos Aires Spanish*. Qualifying paper, Department of Linguistics, New York University.

Roncarati, Claudia, and Jussara Abraçado (eds.). *Português Brasileiro: contato lingüístico, heterogeneidade e história*. Rio de Janeiro: 7Letras.

Scherre, Maria Marta Pereira. 1978. "A regra de concordância de número no sintagma nominal em português". Master's thesis, Pontifícia Universidade Católica-RJ.

Scherre, Maria Marta Pereira. 1988. "Reanálise da concordância nominal em português". Doctoral dissertation, Universidade Federal do Rio de Janeiro.

Schneider, Edgar. 1990. "The Cline of Creole-ness in English-Oriented Creoles and Semi-Creoles of the Caribbean." *English World-Wide* 11: 79–113. https://doi.org/10.1075/eww.11.1.07sch

Schwegler, Armin. 1991. "Negation in Palenquero: Synchrony." *Journal of Pidgin and Creole Languages* 6: 165–214. https://doi.org/10.1075/jpcl.6.2.02sch

Schwegler, Armin. 2002. "On the (African) Origins of Palenquero Subject Pronouns." In *Creoles, their Substrates, and Language Typology*, 225–249. Amsterdam and Philadelphia: John Benjamins.

Schwegler, Armin. 2011. "Palenque(ro): The Search for its African Substrate: Subject Pronouns." *Diachronica* 19: 273–332. https://doi.org/10.1075/dia.19.2.03sch

Singler, John. 1988. " The Homogeneity of the Substrate as a Factor in Pidgin/Creole Genesis." *Language* 64 (1): 27–51. https://doi.org/10.2307/414784

Terrell, Tracy. 1979. "Final /s/ in Cuban Spanish." *Hispania* 66 (2): 599–612. https://doi.org/10.2307/340142

Trudgill, Peter. 2004. *New Dialect Formation: The Inevitability of Colonial Englishes*. Oxford: Oxford University Press.

van Coetsem, Frans. 1988. *Loan Phonology and the Two Transfer Types in Language Contact*. Dordrecht and Providence: Foris. https://doi.org/10.1515/9783110884869

Vieira, Sílvia Rodrigues, and Aline Maria Bazenga. to appear. "A Concordância da terceira pessoa do plural: padrões em variedades do português." In *A concordância verbal em variedades do português: A Interface fonética-morfossintaxe*, ed. by S. R. Vieira. Rio de Janeiro: FAPERJ/Usina de Letras.

Zilles, Ana. 2005. "The Development of a New Pronoun: The Linguistic and Social Embedding of a Gente in Brazilian Portuguese." *Language Variation and Change* 17: 19–53. https://doi.org/10.1017/S0954394505050027

Clause typing and Quebec French *-tu*

Daniela Isac & Justin Royer
Concordia University / McGill University

This paper focuses on the Quebec French particle *-tu*, which occurs in interrogatives, exclamatives and declaratives indicating surprise. We propose that *-tu* is the overt instantiation of a polarity (Pol) head that is neither negative nor positive and that *-tu* also carries an epistemic modality (Mod) feature that is keyed to an ignorant Speaker (S) and a knowledgeable Addressee (A).

Keywords: *tu*, Quebec French, Polarity, Modality, epistemic commitment, Speaker, Addressee

1. General properties of *-tu*[1]

1.1 Distribution

There are several contexts that allow *-tu* in Quebec French (QF), although *-tu* is not obligatory in any of these contexts:

1.1.1 *Interrogatives*

For some speakers, interrogative *-tu* is restricted to yes/no questions, and cannot occur in wh-interrogatives (see also Noonan 1992; Vinet 2000):

(1) *Marie est-tu partie*?
Marie is-*tu* left
'Has Marie left?'

(2) **A qui elle téléphone-tu*?
to who she phones-*tu*
'Who is she calling?'

1. There are two lexical items *tu* in Quebec French (QF): the pronominal *tu* – the 2nd person singular form of a personal pronoun, and the particle *-tu*, which is non-pronominal. In this paper we focus on the latter.

https://doi.org/10.1075/rllt.16.10isa

However, Vecchiato 2000: 141 reports that strings like (2) are possible in some varieties of QF.

1.1.2 *Exclamatives*

Not all exclamative clauses allow *-tu*: while yes/no exclamatives allow it, wh-exclamatives are ungrammatical with *-tu*.

(3) *Il est-tu serviable*!
he is-*tu* helpful
'He's so helpful!'

(4) **Comme elle est-tu belle*! (Vinet 2000: 393, Example 17a)
How she is-*tu* beautiful
'How beautiful she is'

Moreover, some speakers also accept *que*/'that' exclamatives as grammatical when they include *-tu*.

(5) *Qu'elle est-tu fine*!
wh.she is-*tu* nice
'What a nice person she is!'

1.1.3 *Declaratives: Narratives indicating surprise*

Tu can also occur in declarative clauses in QF, provided the latter are part of a narrative context and they express surprise at a sudden or unexpected event.

(6) *Fak-là, il part-tu pas à crier*! (Vinet 2000: 382, Example 2c)
fact-that he starts-*tu* NEG at yell.INF
'Suddenly, he starts yelling'

Tu is always accompanied by the negator *pas* in these contexts, but the *pas* that occurs in these examples does not carry the regular negative meaning, and is instead expletive. This is confirmed by the fact that positive polarity items like *plutôt*/'rather' are licensed in this context, in spite of the presence of *pas*, unlike in genuinely negative sentences like (8).

(7) *Fak-là, elle se met-tu pas plutôt à crier*!
fact-that, she REFL sets-*tu* NEG rather at yell.INF
'Suddenly, she starts yelling.'

(8) *Elle se met (*pas) plutôt à crier.*
She REFL sets (NEG) rather at yell.INF
'She rather starts yelling.' (Vinet 2000: 399, Example 30)

1.1.4 *Imperatives*

The literature also mentions an additional context in which *-tu* can occur, i.e. that of mild or attenuated commands (Vinet 2001, 2002; Morin 2017).

(9) *Tu veux-tu me laisser tranquille*!
you want-*tu* me leave peaceful
'Why don't you leave me alone?' (Vinet 2002: 236, Example 1c)

However, although utterances like (9) can be used as commands, they do not show the properties of genuine imperatives, such as compatibility with covert 2nd person subjects, for instance, but rather of interrogatives. Crucially, genuine imperatives are incompatible with *-tu*.

(10) **Ferme-tu la porte*!
close-*tu* the door
'Close the door!'

We will thus assume that imperative clauses do not really license *-tu*.

1.2 Interpretation of *-tu*

Clauses containing *-tu* have particular interpretive properties, related to the speech act situation. This has been mentioned in the literature either in a general way (Bibis & Roberge 2004; Vinet 2000 claim that *-tu* expresses an illocutionary Force feature), or more specifically, by relating the properties of *-tu* to the epistemic commitment of one of the speech act participants (Bertrand 2014; Kayne 2016; Morin 2017). In particular, the claim is that the use of *-tu* requires a knowledgeable Addressee (A).[2]

In our view, not only the epistemic state of A is important for the licensing of *-tu*, but also that of the Speaker (S). In particular, S must show a low degree of epistemic commitment for *-tu* to be licensed, and the epistemic commitment of A must be higher than that of S. These properties will be discussed in more detail in Sections 2.2 and 3.

2. Proposal

We propose that the *-tu* contexts share two properties: (i) their Pol feature is valued as neither negative nor positive, and (ii) these clauses have an epistemic Mod feature that is anchored to both S and A.

In the first part of this section we will discuss the Pol feature of the *-tu* contexts and will show that these are never negative. We will argue that the incompatibility between *-tu* and negation does not entail that the *-tu* contexts are positive. We

2. The previous literature does not discuss this property for all the contexts that license *-tu*: Kayne 2016 only considers interrogatives, Bertrand 2014 focuses on exclamatives, and Morin 2017 discusses only interrogatives and exclamatives.

will propose instead that *-tu* occurs in contexts whose Pol is non-negative and non-positive. In the second part of the section we argue that the relevant Pol value that instantiates *-tu* is selected by a Mod head that is anchored to the speech act participants, i.e. S and A.

2.1 'Tu' and Polarity

The incompatibility of *-tu* contexts with negation has already been noticed in the literature. Vinet (2002) shows that these contexts do not license NPIs even in the presence of the negator *pas*. While we agree with Vinet (2002) that contexts that allow *-tu* are not negative, we think that this property should be tested with strong NPIs, rather than with weak ones like *qui que ce soit*/'anyone' that appear in Vinet's examples. Strong NPIs occur exclusively in negative clauses and can reliably be used for testing the negative polarity of a clause, while weak NPIs can also occur in other downward entailing contexts, that are not necessarily negative. The examples below show that contexts with *-tu* fail to license strong NPIs like *pantoute*/'at all' (see Burnett & Tremblay 2012 on *pantoute*/'at all' as a strong NPI) even in the presence of the negator *pas*.

(11) *Il parlait-tu pas l'anglais (*pantoute)*?
he spoke-*tu* NEG the.English (at.all)
'Didn't he speak English (at all)?'

(12) *Il est-tu pas serviable (*pantoute)*!
he is-*tu* NEG helpful (at.all)
'Isn't he helpful (*at all)!'

(13) *Fak-là, il entend-tu pas (*pantoute) une bardasse terrible*!
Suddenly he hears-*tu* NEG (at.all) a clatter terrible
'Suddenly, doesn't he hear a terrible clatter.'

This shows that *pas* can only be interpreted as expletive in these contexts and that these contexts are incompatible with genuine, regular negation.

In Vinet's (2002) view the incompatibility of *-tu* with regular negation indicates that *-tu* always occurs in positive clauses. However, not all clauses with positive Pol allow *-tu*, as shown in (14).

(14) *Il part-(*tu) demain.*
he leaves-(*tu*) tomorrow
'He's leaving tomorrow.'

This suggests that *-tu* occurs in a subset of contexts valued as positive, but not in all, and thus that a further distinction is needed. We propose that the set of Pol values should be enlarged so as to include a third possible value, i.e. the non-positive,

apart from the negative and the positive. We could then tentatively state our proposal as in (15):

(15) *Tu* is the overt instantiation of a non-positive Polarity feature.

One argument that (15) is on the right track is that the set of contexts that license *-tu* overlap with the contexts that allow expletive negation in QF.[3] The examples in (11)–(13), which are grammatical without the NPI *pantoute*/'at all' confirm that all contexts that license *-tu* are also compatible with the expletive negator *pas*. (15) can straightforwardly account for the fact that *-tu* is incompatible with regular negation, given that negation and *-tu* instantiate values of Pol which cannot co-occur. (15) can also explain the observed 'optionality' of *-tu*: the 'non-positive' is just one possible value for the Pol feature, along with 'positive' and 'negative'.

2.2 'Tu' and modality

Even though (15) has some explanatory power, it does not completely account for the distribution of *-tu*. In particular, it doesn't explain why only certain types of clauses can have a non-positive Pol. It seems that apart from their Pol, the *-tu* contexts must share at least one other feature.

We propose that the relevant feature is Mod. We saw in section 1.2 that the *-tu* contexts all encode a particular epistemic state of the S/A: S has a low epistemic commitment to the truth of the proposition expressed by these contexts, and A's epistemic commitment is presumed by S to be higher than his/hers. It thus seems reasonable to assume that the *-tu* contexts share a Mod projection hosting an epistemic Mod feature that is keyed to S and A. The link between Mod and S/A can be captured by assuming an S phrase (the 'S Commitment Phrase') and an A phrase (the 'Call on the A'), both above CP, as proposed by Heim et al (2014). Under the assumption that the Mod head bears features that match both the features of the S phrase and of the A phrase, the dependency between Mod and S/A can be formalized by Agree and feature checking.

Similarly, the dependency between Mod and *-tu* (an instantiation of Pol, as proposed in (15)) will be captured by assuming that Mod and Pol share features that enter Agree and feature checking.

---

3. Even though expletive negation can occur in other contexts apart from those that license *tu*, crucially, the contexts that license tu are the only matrix contexts that allow the expletive negator. All the other contexts in which expletive negation occurs in Quebec French are embedded contexts

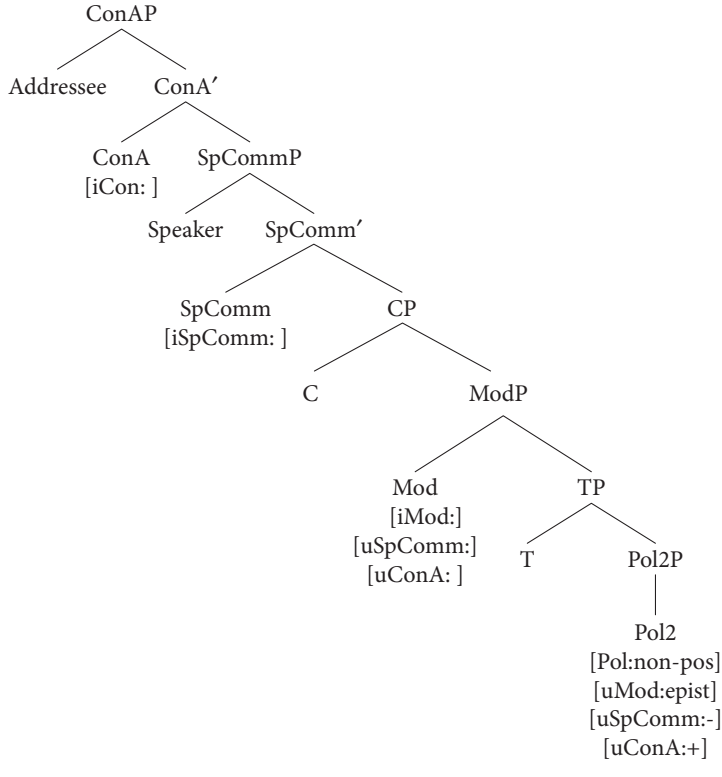
2.3 Analysis

In order to implement the proposal above, we will assume, together with Pesetsky and Torrego (2007), that the interpretable/ uninterpretable distinction is independent from valuation and that both interpretable and uninterpretable features can be valued or unvalued. Moreover, we will assume that Agree results in feature sharing: when Agree applies between a probe feature F at a syntactic location α and a goal feature F at location β, the output is a single feature F shared by two locations.

Given Pesetsky and Torrego's (2007) view on features, the link between Mod and Pol could be captured by assuming that the Mod head has features that are unvalued and thus act as probes, while Pol has matching features that are valued and act as goals. Once Agree takes place, the values of the relevant features are shared in between the Mod head and the Pol head and a link is established. Similarly, the phrases hosting the A and the S will be assumed to have interpretable but unvalued features and the Mod head to carry matching uninterpretable features.

The representation of a *-tu* clause will thus be as in (16):

(16)



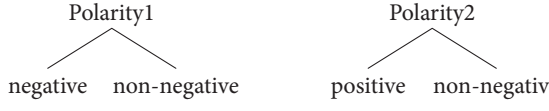
The derivation proceeds as follows. The interpretable but unvalued [iConA:] feature on ConA probes and finds the uninterpretable, unvalued matching [uConA:]

feature on Mod. Agree takes place, establishing a link between the ConA features of these two heads. The resulting shared feature is still unvalued however. Hence [ConA] on the ConA head probes again and enters an Agree relation with the valued [uConA:+] on Pol. Given the pre-established link between ConA and Mod, Agree between [ConA] on the ConA head and [ConA] on Pol results in the valuation of [ConA] in all locations, including Mod. A similar link is established between the [iSpComm:] feature on the SpComm head, and the matching features on Mod and Pol. Finally, the interpretable but unvalued [iMod:] feature on the Mod head probes and finds as its goal the uninterpretable, valued [uMod:epist] feature of Pol, and as a result of Agree this value will be shared among the two heads.

Several observations are in order with respect to the structure in (16):

1. [Mod], [SpComm] and [ConA] cut across the three *-tu* contexts. These features have also been shown to play a role in clause typing and to jointly contribute to the 'signature' of a clause as a particular type (Poletto & Zanuttini 2003; Speas & Tenny 2003, Zanuttini 2008, Isac 2015, among others). We can now explain why the distribution of *-tu* is restricted to only some contexts: *-tu* occurs only in those clauses that contain a ConAP, SpCommP, and a ModP with the feature content shown in (16). Thus, while declarative clauses like (14) could be assumed to have a ModP, as well as a SpCommP, the epistemic commitment of A is irrelevant. Hence it is reasonable to assume that declarative clauses like (14) do not have a ConA phrase at all. That explains why such clauses are incompatible with *-tu*: if *-tu* is merged, its uninterpretable [ConA] feature will remain unchecked and the derivation will crash.
2. In order to explain why the *-tu* contexts can never be negative, we could posit two distinct Pol phrases: one ranging over negative and non-negative values, and one ranging over the positive and non-positive values.

(17)



Each clause type will select either Pol1 or Pol2 and only clauses that select Pol2 allow *-tu*. (15) still holds, but now we can explain why the contexts that license *-tu* can only chose between 'positive' and 'non-positive' as possible values for their polarity feature, and they can never be negative.

3. The distinction in (17) could also account for the optionality of *-tu*. The three contexts that license *-tu* select a Pol2, but the choice between the two possible values of Pol2 (positive or non-positive) is unrestricted. *Tu* is merged only in cases in which the value of Pol is set as 'non-positive'.
4. Given that S/A related phrases are crucial for the *-tu* contexts that such phrases are present in matrix contexts only (Speas & Tenny 2003, Heim et al 2014,

Haegeman & Hill 2013, Isac 2015, among others), we now have an answer as to why *-tu* can only occur in matrix contexts.

3. Accounting for the distributional restrictions

3.1 Interrogatives

Recall that *-tu* is licensed in yes/no questions for all QF speakers, but that *-tu* in wh-interrogatives is accepted only by some speakers. In our view, speakers that do not accept *-tu* in wh-questions analyze this type of clause as containing Pol1, and not Pol2. This is supported by the observation that the S's epistemic commitment in wh-questions is higher than in yes/no questions. In a wh-question the S is presupposing the open proposition containing the wh-variable and is therefore committed to its truth. This is different from S's commitment in a yes/no question, where S has no idea whether the proposition is true or not. This could be implemented by proposing that in wh-questions the SpComm feature must be valued as '+', hence the incompatibility with *-tu*, whose SpComm feature is valued as '–'. The modality of wh-interrogatives has a different flavour than yes/no questions, and hence wh-questions are distinct clause types, ones that select their own type of Pol, i.e. Pol1.

As for speakers that do accept *-tu* in wh-questions, we propose that these speakers analyze wh-questions as clefts. The crucial observation is that speakers that accept wh-interrogatives with *-tu* can insert *que/* 'that' after the wh-item (Vecchiato 2000).

(18) *Où qu'il va-tu*?
where that.he goes-*tu*
'Where is he going?'

This suggests a cleft structure for these clauses, along the lines proposed by Adger and Ramchand (2005) for Scottish Gaelic. This would justify positing a different clause type for these speakers, with specific properties of the functional nodes involved (Mod, SpComm, ConA). In particular, Adger and Ramchand (2005) propose that clefted wh-questions are biclausal: they are copular structures whose semantic predicate is a relative clause, as in (28a). Thus, (18) will have a structure like (19b), with the copula+pronoun cluster phonologically null, just as in Scottish Gaelic.

(19) a. Copula [Wh-phrase] [Relative Clause]
b. ø *[où] [qu'il va-tu]*?
~~It's~~ [where] [that.he goes-*tu*]
'Where is it that he's going?'

In such a structure it is the relative clause that expresses the presupposed proposition to which S is epistemically committed, while the matrix part is strictly speaking free of such presuppositions. Thus, the modality of these wh-questions

is keyed to an S that has a low epistemic commitment, and this in turn licenses a Pol2 phrase (and thus *-tu*), rather than a Pol1.

3.2 Exclamatives

In Section 1.1.2 we showed that *-tu* is grammatical in yes/no exclamatives but not in wh-exclamatives, and moreover that for some speakers *-tu* is also grammatical in exclamatives introduced by *que*/ 'that'.

A number of authors (Zanuttini & Portner 2003; Michaelis & Lambrecht 1996; Rett 2011, etc.), have shown that exclamatives are associated with a variable whose possible values form a scale. Moreover, exclamatives have been shown to give rise to the implicature that the proposition they denote lies at the extreme end of this scale. This implicature arises, among other things, from a quantificational operator merged in the CP domain of exclamatives, according to Zanuttini and Portner 2003. The operator's role is to expand the domain of individuals/properties in the range of the variable. This widening operator is crucial for typing a clause as an exclamative. Clauses that simply express surprise, but do not contain such an operator, are not genuine exclamatives, but simply 'proposition exclamations' (PE) (Rett 2008).

(20) a. Sue wore orange shoes! (Rett 2008: 601. PE)
b. (Boy,) can Robin bake pies! (Rett 2008: 613. E)

(20a) simply expresses that the Speaker finds the proposition that Sue wore orange shoes surprising. (20b) cannot be used to express surprise at the proposition that Robin can bake pies and is used instead to express surprise that Robin can bake a large amount of pies for instance, or that Robin can bake pies that are delicious to a certain (high) degree.

If the two types of yes/no exclamatives are indeed different, we expect this distinction to be reflected in the the way *-tu* is licensed. Indeed, *-tu* can only be used in E's like (21), which express surprise at the degree to which a certain property holds, but not in PE's like (22), which simply expresses surprise that the proposition 'my flight is cancelled' is true.

(21) *A n'en parle-tu des langues*!
she of.them speak-*tu* of.the languages
'How many languages she speaks!'

(22) *Mon vol est-(*tu) annulé*!
My flight is-*tu* cancelled
'My flight is cancelled!'

E's like (21) are compatible with *-tu* in our analysis because they contain a Pol2, which is in turn related to the Mod of the exclamative, and the epistemic states of

S and A. In (21) the S expects the number of languages spoken by the subject to range only up to a particular degree, and the actual value of that number exceeds that expectation. Since S's expectations are exceeded and thus negated, S's epistemic commitment is low, since the proposition is unexpected for S. The A, on the other hand, is called upon to confirm that the actual value for the variable contained in the exclamative is higher on the scale than expected by S (as shown by Bertrand 2014). In other words, A is asked to confirm that the number of languages the subject speaks goes beyond the range expected by S. In sum, yes/no exclamatives show similar clause typing ingredients as yes/no questions: epistemic Mod keyed to an ignorant S and a knowledgeable A. This explains why yes/no exclamatives select a Pol2 and thus license *-tu*. Let us now move on to explaining why *-tu* does not occur in wh-exclamatives. One reason for this restriction is that for some speakers at least wh-exclamatives are not grammatical at all in QF, unless these exclamatives are nominal and lack a verb altogether, as in (23).

(23) *Quelle belle maison*!
what beautiful house
'What a beautiful house!'

For these speakers, *-tu* is ungrammatical in these contexts because they lack a Tense projection, and hence a Pol projection.

For other speakers, though, at least some of the clausal wh-exclamatives are grammatical, as shown in (24) from Dubois (2000: 130), so the question is why do these speakers reject *-tu* in these contexts.

(24) *Quels tableaux extraordinaires j'ai vu là*!
what paintings extraordinary I.have seen there
'What extraordinary paintings I saw there!'

We propose that these speakers analyze wh-exclamatives like (24) on a par with wh-questions. As argued by Michaelis and Lambrecht (1996), Zanuttini and Portner (2003), Abels (2010), wh-exclamatives are similar to wh-interrogatives in that they denote sets of propositions whose content is presupposed in the context of utterance. The S is therefore not ignorant in these contexts. This is particularly apparent in (24) where the wh-item is a determiner and the restriction of the wh-variable is explicitly provided by the NP following the determiner. S knows the domain over which the wh-variable ranges: the set of paintings in (24). Given that *-tu* occurs only in clauses whose Mod is keyed to an ignorant S, *-tu* is not licensed in (24).

The exceptional status of *que* exclamatives can be accounted for by assuming that *que* does not actually have wh-features for speakers who accept strings like (5) and therefore that *que* exclamatives are similar to yes/no rather than

wh-exclamatives for these speakers. Even though we will not offer an analysis of *que* exclamatives, it seems reasonable to assume that the *que* in these contexts is the overt expression of a degree operator, rather than a genuine wh-item.

3.3 Declaratives

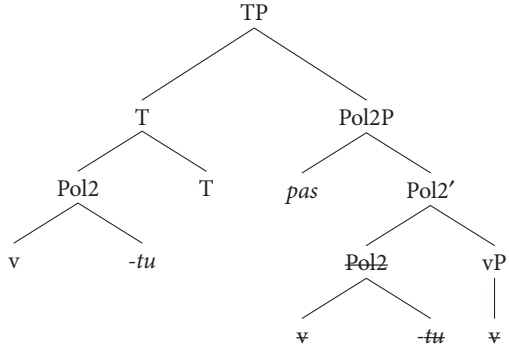
Finally, the distribution of *-tu* in declaratives can also be accounted for by our proposal. Recall that not all declaratives license *-tu*, as mentioned in section 1. This suggests that declarative clauses expressing surprise have different properties than regular declaratives and perhaps fall into a different clause type category altogether. This is supported by the fact that the interpretation of these clauses is closer to exclamatives than to regular declaratives. These contexts necessarily include an adverb expressing surprise, which acts as a mirative marker. Like all mirative markers, this adverb introduces a partition between the set of propositions that are expected and the set of propositions that are not expected by S, and are thus surprising (DeLancey 2001). Thus, just as in the case of exclamatives, S's epistemic commitment to the expressed proposition is low, since the proposition is unexpected for S. In the example below, the narrative context for the underlined sentence that contains *-tu* explicitly states that S expected the horses to have gone, and the sentence including *-tu* expresses surprise at the unexpected fact that the horses reappeared.

(25) *Marche, marche, marche… pas de nouvelles des chevaux. J'ai fini par penser que j'avais vraiment rêvé pis j'me sus mis à siffler pour oublier toute ça.*
'Walk, walk, walk … no sign of the horses. I ended up thinking it had all been a dream and I started whistling in order to forget everything.'
Tout d'un coup j'entends-tu pas une bardasse terrible en arrière
All of.a sudden I.hear-*tu* neg a noise terrible n behind
d'la montagne.
of.the mountain.
'All of a sudden I hear a terrible noise coming from behind the mountain.'
[…] Pis… j'ai vu …les huit grands chevaux sortir d'en
'[…] Then… I saw them… the eight big horses coming
arrière de la montagne … (Tremblay 1978: 251)
from behind the mountain …'

The low commitment of S is directly linked to the fact that these clauses express an unexpected proposition. Other declaratives, that lack this property, do not license *-tu*. Regular declaratives assert a proposition which is added to the common ground and S is highly committed to the truth of this proposition. The Mod of regular declaratives therefore shows properties that differ from the Mod of clauses indicating surprise.

Last, but not least, our analysis can also account for the order of *-tu* and *pas* in these contexts. Recall that in declarative clauses expressing surprise *-tu* must co-occur with *pas*, and moreover *-tu* immediately precedes *pas*. While we will have nothing to say about the co-dependency between *-tu* and *pas*, according to our analysis the co-occurrence of *-tu* and *pas* is predicted to be possible only if *pas* is not genuine negation. This prediction is borne out since, as shown in Section 2.1, *pas* cannot license strong NPIs in these contexts. The relative order of *-tu* and *pas* can be easily accounted for under the assumptions that (i) *-tu* and *pas* occupy the head and the Spec positions of the PolP, respectively; and (ii) the verb raises to at least as high as TP in QF.

(26)



Given the general constraints on head movement, the verb first raises to Pol, where it left-adjoins to *-tu*, and the resulting [V-*tu*] head further moves to the Mod head, thus coming to precede *pas*.

3.4 Imperatives

Finally, imperatives do not license *-tu* because this clause type selects a Pol1 type of Pol feature. This is confirmed by the fact that the negator that occurs in imperatives is a regular negator, that does license strong NPIs.

(27) *Bouge pas pantoute*!
move NEG at.all
'Don't move at all!'

Moreover, the Mod of imperatives is deontic, and not epistemic. Rather than showing any epistemic commitment to the truth of the proposition expressed by the imperative, S is the source of an instruction to A to add the proposition on her 'TO-DO' list (Portner 2004). Since in our proposal only epistemic modality can select a Pol2, imperatives will never do so.

4. Previous analyses

Most previous analyses on *-tu* (Kayne 2016; Rowlett 2007; Elsig 2009; Pollock 2006; Bibis & Roberge 2004; Noonan 1992) focus exclusively on interrogatives, with the exception of Vinet (2000, 2002) and Morin (2017). Vinet (2000, 2002) proposes that *-tu* is an operator with an emphatic positive reading. While this is intuitively compatible with exclamatives and surprise declaratives, it is not clear how this view can be extended to interrogatives. Vinet (2002) claims that the emphatic meaning of *-tu* in questions is manifested in *-tu* bearing focal stress in these contexts, which in turn indicates that an affirmative answer is expected. However, no reliable test is proposed for determining whether a yes/no question with *-tu* expects a positive answer. Moreover, if *-tu* is an operator with an emphatic positive reading, one cannot rule out *-tu* in regular declaratives since there is no reason why these clauses cannot be emphatically positive.

In Morin's (2017) view *-tu* is an interrogative particle that heads a syncretic projection FP with features of both I and C.

(28) [FP NP_{subj} [F' [F [V *aime*][F *-tu*]]… [VP t_{NP} t_{V} *les chats*]]]

Apart from the fact that Morin (2017) only discusses the occurrence of *-tu* in interrogatives and exclamatives, while our analysis extends to declaratives of surprise as well, our analysis has one point of similarity with Morin's (2017). In order to account for the incompatibility between *-tu* and negation, Morin (2017) assumes that *-tu* carries a Pol feature. The complementarity between *-tu* and the negative marker is explained by the fact that both of them 'satisfy' a Pol feature and they therefore compete. The intuition behind Morin's (2017) proposal is clearly similar to our view on *-tu*. However, our analysis differs in several respects. First, for Morin (2017) *-tu* carries both a Pol feature and an interrogative feature. While this proposal can be extended to exclamatives by assuming that the latter have the same syntax as interrogatives, it cannot be carried over to declarative clauses expressing surprise. It is not clear why such clauses would have an interrogative feature. Second, in Morin's (2017) analysis the [Pol] feature can be checked either by the negative head, or by *-tu*. This explains indeed the fact that the negative head and *-tu* cannot co-occur, but it also leads to the expectation that *-tu* contexts would either contain a Neg head or would license *-tu*, since *-tu* and the Neg head exhaust the possible values of [Pol] in her analysis. However, this expectation is not borne out; all of the contexts that license *-tu* can occur without *-tu* and without negation.

5. Conclusions

In this paper we argued that QF *-tu* is the overt expression of a Pol head valued as non-positive. We proposed a distinction between two types of Pol phrases: Pol1, that ranges over the negative/non-negative values, and Pol2, that ranges over positive/non-positive values and that the *-tu* contexts select Pol2. Thus, in our system, the non-positive value alternates with the positive only. This explains why the contexts that allow *-tu* can only be interpreted as non-positive (in the presence of *-tu*) or positive (when *-tu* is absent), but never negative. Apart from their Pol, what sets the contexts that allow *-tu* apart in our view is their particular type of Mod, as well as the epistemic commitment of the speech act participants: all the *-tu* contexts share an epistemic modality that is anchored to an ignorant S and a knowledgeable A. Our analysis thus supports the view that various clause types are identified not by one unique feature, like Force, but by multiple features, such as Mod, and S/A related features.

References

Abels, Klaus. 2010. "Factivity in Exclamatives is a Presupposition." *Studia Linguistica* 64: 141–157. https://doi.org/10.1111/j.1467-9582.2010.01164.x

Adger, David, and Gillian Ramchand. 2005. "Merge vs. Move: Wh-Dependencies Revisited." *Linguistic Inquiry* 36 (2): 161–193. https://doi.org/10.1162/0024389053710729

Bertrand, Anne. 2014. *Exclamatives en -tu, donc et assez en français québécois*. Montreal, Quebec: Université de Montréal dissertation.

Bibis, Nick, and Yves Roberge. 2004. "Marginal Clitics." *Lingua* 114 (8): 1015–1034. https://doi.org/10.1016/S0024-3841(03)00103-7

Burnett, Heather, and Mireille Tremblay. 2012. "An Extra-Strong NPI? *Pantoute* in Quebec French." *UCLA WPL* 17: 1–8.

Dubois, Caroline 2000. *La Grammaire de l'exclamation: Aspects théoriques, français de référence et français québécois*. Sherbrooke, Quebec: Université de Sherbrooke MA Thesis.

Elsig, Martin. 2009. *Grammatical Variation Across Space and Time: The French Interrogative System*. Amsterdam: John Benjamins. https://doi.org/10.1075/silv.3

Heim, Johannes, Hermann Keupdjio, Zoe Wai-ManLam, Adriana Osa-Gómez, and Martina Wiltschko. 2014. "How to Do Things with Particles." *Proceedings of the CLA*, 1–15.

Isac, Daniela. 2015. *The Morphosyntax of Imperatives*. OUP. https://doi.org/10.1093/acprof:oso/9780198733263.001.0001

Kayne, Richard. 2016. "The Silence of Heads." *Studies in Chinese Linguistics* 37 (1): 1–37. https://doi.org/10.1515/scl-2016-0001

DeLancey, Scott. 2001. "The Mirative and Evidentiality." *Journal of Pragmatics* 33 (3): 369–382. https://doi.org/10.1016/S0378-2166(01)80001-1

Michaelis, Laura A., and Knud Lambrecht. 1996. "Toward a Construction-Based Model of Language Function: The Case of Nominal Extraposition." *Language* 72: 215–247. https://doi.org/10.2307/416650

Morin, Annick. 2017. *Questioning Particles: A Cross-Linguistic Approach to Quebec French Polar Interrogatives*. Toronto: University of Toronto dissertation.

Noonan, Máire, B. 1992. *Case and Syntactic Geometry*. Montreal, Quebec: McGill University dissertation.

Pesetsky, David, and Esther Torrego. 2007. "The Syntax of Valuation and the Interpretability of Features." In *Phrasal and Clausal Architecture. Syntactic Derivation and Interpretation. Linguistik Aktuell/Linguistics Today* 101, ed. by S. Karimi, V. Samiian, and W. K. Wilkins, 262–294. Amsterdam: John Benjamins.

Poletto, Cecilia, and Raffaella Zanuttini. 2003. "Making Imperatives: Evidence from Central Rhaetoromance." In *The Syntax of Italian Dialects*, ed. by C. Tortora, 175–207. Oxford University Press.

Pollock, Jean-Yves. 2006. "Subject Clitics and Complex Inversion." In *The Blackwell Companion to Syntax*, ed. by Martin Everaert, and Henk van Riemsdijk, 601–659. Malden, MA: Blackwell Publishing.

Portner, Paul. 2004. "The Semantics of Imperatives Within a Theory of Clause Types." *Semantics and Linguistic Theory* 14: 235–252. https://doi.org/10.3765/salt.v14i0.2907

Rett, Jessica. 2011. "Exclamatives, Degrees and Speech Acts." *Linguistics and Philosophy* 34 (5): 411–442. https://doi.org/10.1007/s10988-011-9103-8

Rett, Jessica. 2008. "A Degree Account of Exclamatives." In *Proceedings of SALT* XVIII, ed. by J. Rett, T. Friedman, and S. Ito, 601–618. https://doi.org/10.3765/salt.v18i0.2470

Rowlett, Paul. 2007. *The Syntax of French*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511618642

Speas, Peggy, and Carol Tenny. 2003. "Configurational Properties of Point of View Roles." In *Asymmetry in Grammar* 1, ed. by A. M. Di Sciullo, 315–345. https://doi.org/10.1075/la.57.15spe

Tremblay, Michel. 1978. *La grosse femme d'à côté est enceinte*. Leméac.

Vecchiato, Sara. 2000. "The *ti/tu* Interrogative Morpheme in Quebec French." *Generative Grammar in Geneva* 1: 141–163.

Vinet, Marie-Thérèse. 2000. "Feature Representation and *-tu (pas)* in Quebec French." *Studia Linguistica* 54 (3): 381–411. https://doi.org/10.1111/1467-9582.00071

Vinet, Marie-Thérèse. 2001. *D'un français à l'autre: La syntaxe de la microvariation*. Montreal: Fides.

Vinet, Marie-Thérèse. 2002. "*Tu* as a (Super) Positive Marker in Quebec French." In *Romance Languages and Linguistic Theory (Selected Papers from 'Going Romance' Groningen, November 2002)*, ed. by Reineke Bok-Bennema et al., 235–252. Amsterdam/Philadelphia: John Benjamins.

Zanuttini, Raffaella, and Paul Portner. 2003. "Exclamative Clauses: At the Syntax-Semantics Interface." *Language* 79 (1): 39–81. https://doi.org/10.1353/lan.2003.0105

Zanuttini, Rafaella. 2008. "Encoding the Addressee in Syntax: Evidence from English Imperative Subjects." *Natural Language and Linguistic Theory* 26 (1): 185–218. https://doi.org/10.1007/s11049-007-9029-6

The syntax of superlative phrases in Romance

Nicoletta Loccioni
UCLA

In this paper, prenominal and postnominal superlatives in Romance are argued not to differ as a result of NP-movement (or lack thereof) to cross the adjective, but to result from two different derivations. In prenominal superlatives, the adjective is attracted to a high scope position in the DP-area of numerals where it can't be bypassed by the NP. In postnominal superlatives, the superlative starts out as a predicate (a maximalizing relative clause over degrees) and it ends up postnominal (as any other relative clause in Romance).

Keywords: superlatives, French, Italian, adnominal modification, DP structure, degree phrases, definiteness

1. Introduction

Like ordinary forms, (some) superlative adjectives can be prenominal or postnominal in Romance (exemplified by Italian below):

(1) a. la bella ragazza
the beautiful girl

b. la ragazza bella
the girl beautiful
'the beautiful girl'

(2) a. la più bella ragazza
the more beautiful girl

b. la ragazza più bella
the girl more beautiful
'the most beautiful girl'

Despite this first similarity, the properties of superlative forms w.r.t. their adnominal position do not parallel the ones of ordinary forms in several ways:

a. unlike ordinary adjectives, prenominal superlatives do not need to have the unambiguous 'direct-modification' interpretation advocated by Cinque 2010

https://doi.org/10.1075/rllt.16.11loc

and can have a 'indirect-modification interpretation' (this is based on new observations that I will discuss below);

b. superlative forms always license ellipsis, even when the ordinary form of the adjective does not. Here exemplified by the adjective *arrabbiato* 'angry'.

(3) a. *L' arrabbiato
the angry
int. 'the angry one'

b. Il più arrabbiato
the more angry
'the angriest (one)'

c. unlike ordinary adjectives, French postnominal superlatives require a second overt determiner (see Kayne 2008 a.o.):

(4) la fille (*la) belle
the girl the beautiful
'the beautiful girl'

(5) la fille *(la) plus belle
the. girl the more beautiful
'the most beautiful girl'

In this paper, I will account for these differences arguing that the prenominal and the postnominal positions of superlative forms in Romance correspond to two independent strategies to obtain superlative import. I will refer to them as *attributive* and *predicative* constructions respectively.

In the case of prenominal superlatives, the adjective is attracted to a high scope position in the left periphery of the DP where it can't be bypassed by the NP. This higher position available to superlative forms is responsible for the properties outlined in (a) and (b) above. Prenominal superlative phrases can be broken down into an ordinal-like element (responsible for the left-peripherical position), a comparative and a nominal phrase. Semantically, this results in the ordinal picking out an individual out of an ordered set. Informally, the phrase in (2a) gets to mean something like *the first in a class of girls ranked w.r.t. their beauty*.

Postnominal superlatives on the other hand are predicates, which means that they do not embed a nominal projection. Specifically they are maximalizing relative clauses over degrees, which refer to the highest degree of a property in a class. I will show that this MAX operator is spelled out in French as a mandatory definite determiner and it is left unpronounced in Italian (or Spanish). (2b) can then be paraphrased as the *unique individual in a class of girls who is associated with the maximal degree of beauty*.

In most cases, prenominal superlatives and postnominal ones end up being truth-conditionally equivalent but I will show that the nature of the attributive superlative limits the range of interpretations that it can yield. So-called relative readings are in fact only compatible with postnominal superlatives.

2. Prenominal superlatives

The dual source model of adnominal modification presented by Cinque 2010 does not predict the possibility for indirect modification adjectives to appear prenominally in Romance, unless additional assumptions are made. Nevertheless, superlative forms of indirect modification adjectives can end up prenominally in Italian. In this section, I will first briefly review Cinque's (2010) dual source model and then I will show that a movement analysis is necessary in order to derive the apparently anomalous behavior of superlatives.

2.1 Cinque's (2010) dual source model

According to the dual source model presented by Cinque 2010, adnominal modifiers can enter the derivation in two independent ways: either as direct modifiers or as indirect modifiers. In the first case, they are analyzed as phrasal specifiers of dedicated functional heads of the extended projection of the noun. They are structurally closer to the noun and are semantically associated with individual-level, nonrestrictive, absolute readings. Indirect modification adjectives on the other hand are merged above the functional projections hosting direct modification adjectives. They are predicates of reduced relative clauses and are semantically associated with intersective, relative and stage-level readings (but they are also compatible with individual-level interpretations).

The order of Merge is one where modifiers are rigidly ordered w.r.t. each other: indirect modifiers are always merged higher than direct ones:

(6) Order of Merge:
[D …[AP_{ind} … [AP_{ind} [AP_{dir}… [AP_{dir}… NP]]]]]

According to Cinque 2010, different conditions on NP movement and extraposition derive the surface orders for Romance and Germanic in (7) and generate the following one-way implications: only direct modification adjectives are able to appear prenominally in Romance, whereas only indirect modifiers can be postnominal in English.

(7) a. Italian: [D … AP_{dir} NP AP_{dir} AP_{ind}]
b. English: [D … AP_{ind} AP_{dir} NP AP_{indir}]

(8) a. [ROMANCE] If prenominal → direct modification source.
b. [ENGLISH] If postnominal → indirect modification source.

Importantly, the generalization in (8a) does not hold in the case of superlatives: unlike the ordinary form of 'povero', which can only mean 'pitiable' prenominally (as shown in (9)), the prenominal superlative form has the indirect modification interpretation 'poor' (as shown in (10)):

(9) Una povera ragazza
'A pitiable girl'
#'A poor girl'

(10) Liverpool […] dall' 84 è tra le ***più povere*** città europee
Liverpool from '84 is among the more povere cities European
'Since 1984, Liverpool is among the poorest European cities'
(from www.repubblica.it)

2.2 A movement analysis of prenominal superlatives

In the previous section, I showed that superlatives challenge Cinque's generalization about the unambiguousness of the prenominal position in Romance in that superlative forms of indirect modification adjectives can appear prenominally even if the ordinary forms cannot. I interpret this fact as evidence that superlative adjectives can appear higher than ordinary ones.[1] Two questions then arise:

i. what is the position they are associated with?
ii. why can they appear higher than ordinary form?

As for the first question, I adapt (and expand) the order of Merge argued for by Cinque 2005 and I argue that attributive superlatives are associated with the position(s) in (11):

(11) Order of Merge:
a. [D… [Num$_{ord}$… [Num$_{card}$ … **Sup** [(RC) [Cl… [A… NP]]]]
b. [D… [Num$_{ord}$… **Sup** [Num$_{card}$ … [(RC) [Cl… [A… NP]]]]

They are outside the extended projection of the NP, above all the modifiers. They are also structurally higher than relative clauses, as suggested by the fact that they can license NPIs and subjunctive mood inside relative clauses, as shown in (12):

1. This idea is far from being new in the literature. Kayne 2008 reports the fact that in Persian superlatives end up prenominal (while ordinary adjectives and comparatives are generally postnominal) and for Romance superlatives – and superlatives only – posits a 'superlative preposing' operation.

(12) la più bella ragazza che io abbia mai visto
the COMP beautiful girl that I have.SUBJ ever seen
'the most beautiful girl I've ever seen'

Cardinal numbers can either follow or precede prenominal superlatives. The different position of the numeral w.r.t. to the superlative has an effect on the interpretation. In particular, when a cardinal number follows the superlative expression in (13b), presentations are paired up and pairs of presentations of some length are compared. When the cardinal number precedes the superlative (as in (13a)), individual presentations are instead compared.

(13) a. le due più lunghe presentazioni CARD > SUP
the two more long presentations
b. le più lunghe due presentazioni SUP > CARD
the more long two presentations

The cooccurence of superlative and ordinal number is an innovation in standard Italian that not every speaker accepts (see Grasso 2007). For the speakers who accept it, the natural order is the same as English with the superlative following the ordinal number.

(14) la seconda più alta montagna che sia mai stata scalata
the second more high mountain that is.SUBJ never been climbed
'the second highest mountain that was ever climbed' ORD > SUP

I turn next to the second question, why do superlative forms end up in this left peripherical position that I have identified with the region of numerals? I am going to suggest that this results from their *quantificational component* that can attract the adjective to a scopal position outside the extended nominal projection, which is the constituent that provides the comparison class.[2]

---

2. A potential problem for the claim that the superlative appears in its scope position is created by prenominal possessives. They precede the superlative, even though they clearly participate in the calculation of the comparison class. In (i), for instance the superlative compares 'mistakes of mine' and not simply 'mistakes' despite the fact that the superlative follows the possessive.

(i) il mio più grande sbaglio
the my COMP big mistake
'my biggest mistake'

This tension between the position of prenominal possessives and their interpretative properties can be easily resorved once we assume that possessives end up prenominal as a result of movement from a lower position, as suggested by Cardinaletti 1998. This would allow to

In the case of Romance, superlative forms are not morphologically distinguished from comparative forms. Despite that, I take the high position of prenominal superlatives to be informative about the internal make up of superlatives. In particular, it suggests that an ordinal-like element is active in Romance even if it is not morphologically overt. I label it SUP and I assume it to be merged in this left-peripherical position:

(15) $[_{DP}$ the [SUP $[_{FP}$ [AP $[_{NP}$]]]]

Nothing new has to be said about how the adjectival phrase enters the derivation. In particular other syntactic material can be merged between the (comparative form of the) adjective and SUP. Relative clauses are examples of such interveners:

(16) $[_{DP}$ the [SUP $[_{FP}$ $[_{CP}$ (relative clause)] [AP $[_{NP}$]]]]

The position of the superlative form of the adjective is then the result of movement of the AP, which is attracted to the position of SUP.

2.3 Elliptical superlatives as attributive superlatives

Cinque 2010 (p. 50) discussed the fact that only (some) direct modifiers (as (17c)) are possible in elliptical DPs introduced by the definite article in Italian, as opposed to indirect modifiers (17a) and full relative clauses (17b).

(17) a. *Le { arrabbiate/ orgogliose dei propri figli}
the angry proud of their children
(understood: *madri*, 'mothers')

b. *Le che sono state pubblicate
the that have been published
(understood: *riviste*, 'journals')

c. Le altre/ precedenti/ principali/ probabili
the other former main probable
(understood: *conseguenze*, 'consequences')

But superlative forms of indirect modification adjectives license ellipsis:

(18) La più { arrabbiata/ orgogliosa di suo figlio }
the more angry proud of her son
(understood: *madre*, 'mother')
'the angriest/ proudest of his son'

---

interpret the possessive in the base (pre-movement) position, in the extended projection of the NP and within the scope of the superlative. This is shown in (ii).

(ii) $[_{DP}$ *il* $[_{XP}$ *mio*$_i$ … $[_{ZP}$*più grande* $[_{YP}$ *sbaglio*$_k$ $[_{NP}$ t$_i$ [t$_k$ …

This may suggest that elliptical superlatives in Romance are attributive structures, as high as prenominal superlatives. It was originally suggested by Sleeman 1996 that the elliptical superlatives like (18) are attributive structures that license the empty noun because of their partitive meaning:

(19) Preferisco il più grande *pro*
I.prefer the more big *pro*
'I prefer the biggest' Sleeman 1996

If correct, this would imply that when the NP is omitted, the adjective can move in a way that is not possible otherwise. The adjective *arrabbiata* 'angry.F' can never be prenominal, not even when superlative, as shown in (20). Despite that, it is able to license ellipsis (see (18)).

(20) .?/*la più arrabbiata madre
the more angry mother

This correlation between the presence of an overt noun and freedom of movement of the adjective is mysterious, yet not unprecedented (see Rohena-Madrazo 2007 for similar considerations about Puerto Rican Spanish superlatives).

2.4 Attributive superlatives are individual-based superlatives

Descriptively speaking, attributive superlatives in Italian are individual-based superlatives. They cannot associate with focus and yield relative readings (on absolute vs. relative readings see Szabolcsi 1986; Heim 1985; Heim 1999; Sharvit & Stateva 2002 a.o).

Since the object position of *relation have* can host relative readings of superlatives (given their semantic indefiniteness) but not absolute ones, (21a) is interpreted as *John has a prettiest sister than anybody does*. The same relative reading is not available in Italian attributive superlatives and (21b) and (21c) are deviant, as a result[3]

(21) a. JOHN$_F$ has the prettiest sister
b. #La più bella sorella, ce l' ha Luigi.
The more beautiful sister, CE CL has Luigi
int. 'LUIGI has the most beautiful sister'
c. #La più bella, ce l' ha Luigi.
The more beautiful, CE CL has Luigi
int.'LUIGI has the most beautiful (one)'

3. A clitic left dislocated construction is used here to allow the subject to be in a post-verbal focus position. Since focus is known to facilitate relative readings, the example was constructed this way to make the relative reading the most prominent.

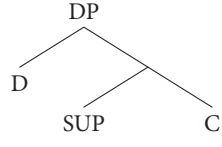
I take this to suggest that what attributive superlatives compare is individuals, as opposed to degrees. A sketch of a very informal and oversimplified compositional account for the phrase (22) is given below.

(22) il più grosso gatto bianco
the more big cat white
'the biggest white cat'

First, the comparison class is roughly determined by the extended nominal projection, [$_{FP}$ gatto bianco]. This means that only (relevant) white cats are compared. Second, the comparative morpheme scopes above the NP (and below SUP) and returns an ordered list. That is, white cats are ordered based on their size. Then, SUP functions as an ordinal and takes the first ranked element. Finally, the definite determiner tests for uniqueness and returns the unique first-ranked cat, if there is one.

2.5 Summary

To sum up, in this section I showed that when superlative, adjectives can appear higher than ordinary forms. In particular, they can be associated with a high attributive position in the DP that cannot be bypassed by the NP in Romance. I argued that they appear in this high position as a result of movement to the merger position of SUP, which is in the left periphery of the DP as shown in (11). I also claimed that the position of SUP determines how the comparison class for the superlative is calculated. In fact, its sister **C** denotes a set of individuals ordered with respect to a given property (the one denoted by the adjective) and SUP picks the first out of that set.



Only a subclass of adjectives can be attracted to this high position when the noun is overt. For reasons that remain to be understood, this class is slightly larger than the class of adjectives that can end up prenominally in the ordinary form. More work needs to be done to properly isolate the properties of the members of this class.

Lastly, I followed Sleeman 1996 in claiming that elliptical superlatives are attributive superlatives. This would imply that when the noun is not overt virtually any adjective can be attracted by the superlative morpheme to this high position.

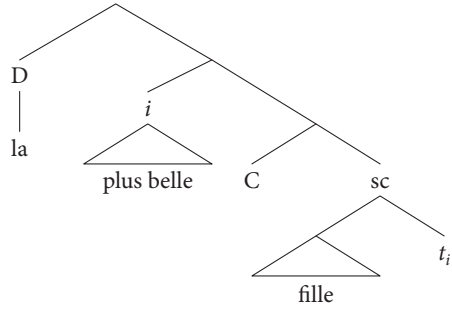
3. Postnominal superlatives

In this section I turn to postnominal superlatives as in (2b) and I will argue that they are predicates (reduced relative clauses) and that the mandatory second definite determiner in French follows from a language-specific fact about how definiteness of DegPs is expressed. In order to do so, I will call into question the analysis put forth by Kayne 2008; and the claim made by Matushansky 2008 that superlatives cannot function as predicates. I will briefly review Kayne's (2008) analysis before presenting my own in Section 3.2. In Section 3.3, I will then discuss the problems of an alternative proposal, namely that postnominal superlatives are attributive superlatives with a null noun.

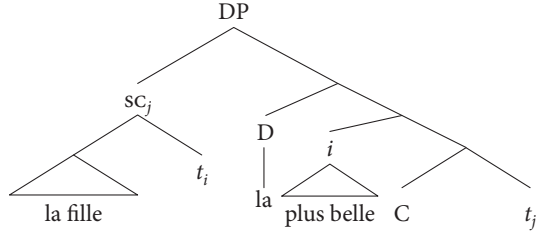
3.1 Kayne's (2008) analysis

Kayne 2008 argues for the following two structures for prenominal and postnominal superlatives in Romance:

(23) La plus belle fille



(24) La fille la plus belle



Here are the core properties of Kayne's (2008) proposal. First, an operation of 'superlative preposing' is responsible for the movement of *plus belle* in both structures to the same position. Second, prenominal superlative phrases differ from postnominal ones in two ways: (i) the subject of the small clause lacks a determiner

and (ii) small clause preposing does not occur. Lastly, the difference between Italian and French, repeated below, is explained based on the parametric difference in (26):

(25) a. la ragazza (*la) più bella
b. la fille *(la) plus belle
the.F.SG. girl the.F.SG. more beautiful.F.SG.
'the most beautiful girl'

(26) a. In Italian, a definite D with a filled Spec can and must be unpronounced.
b. In French, a filled Spec does not license non-pronunciation for a definite D.

Thus, the structure for Italian (25a) would be identical to the one in (24), with the only difference that the filled specifier of the DP in Italian would cause the determiner to be unpronounced.

3.2 My current proposal: Postnominal superlatives as predicates

The main problem with Kayne's analysis that I would like to point out is that it fails to extend to other DP-internal superlatives, where the same asymmetry is found w.r.t. the pronunciation vs. non-pronunciation of D. This is shown in the case of full relative clauses in (27), as well as adverbial (as in (28)) and quantity superlatives (as in (29b)).[4] All the (a)-examples are from French, whereas all the (b)-examples are from Italian.[5]

(27) a. Marie est celle qui est *(la) plus énervée
M. is the.one who is the more angry
b. Maria è quella che è più arrabbiata
M. is the.one who is more angry
'Marie is the one who is the angriest'

(28) a. La secrétaire qui écrit *(le) mieux
the secretary who writes the better
'the secretary who writes the best'
b. la segretaria che scrive (*il) meglio.
the secretary who writes the better
'the secretary who writes the best'

4. (28a) and (29a) are ungrammatical when the second definite determiner is omitted only under a superlative interpretation.

5. Other Romance languages such as Spanish and Catalan as well as some Middle French dialects pattern with Italian.

(29) a. La fille qui a *(le) plus d' argent
the girl who has the more of money
'the girl who has the most money'

b. la ballerina che ha (*i) più soldi
the dancer who has more money
'the dancer who has the most money'

DP-internally, the difference between French and Italian-type languages can be stated in the following way: French must have a dedicated second overt determiner preceding *plus*, whereas Italian cannot have it. If we move away from superlatives embedded inside definite DPs and we look at the sentential level, we find the following facts: in superlatives that have relative readings only,[6] superlatives cannot be formed in Italian simply by adding a definite determiner to a comparative. To see this, compare the ungrammatical Italian examples in (30a) and (31b) to their French counterparts in (30a) and (31a).

(30) a. Marie écrit le mieux
Maria writes the better
'Maria writes the best'

b. *Maria scrive il meglio
Maria writes the better
Int. 'Maria writes the best'

(31) a. Marie a le plus d' amis
Maria has the more of friends
'Maria has the most friends'

b. *Maria ha i più amici
Maria has the more friends
Int. 'Maria has the most friends'

The data in (27) through (31) indicate that the two (types of) languages only differ in the pronunciation of a dedicated determiner inside the predicate and not in the pronunciation of the main determiner in the spine of the DP, as argued by Kayne 2008 (see (24)). It is necessary then to make a distinction (much in the spirit of Krasikova 2012) between two different functions that can be played by the definite article: it can reflect the definiteness either at the DP-level or at the DegP-level:[7]

(32) a. $[_{DP}$ **D** $[\ [_{NP} \ldots N\ ]\ ]\ ]$
b. $[_{DegP}$ **D** -est DEGREE $]$

6. For discussion of the difference between *absolute* and *relative* readings of superlatives, see Heim 1985, Szabolcsi 1986, Sharvit and Stateva 2002 among many others.

7. Here I use the functional noun DEGREE to refer to AMOUNT or NUMBER as well.

The variation we discussed above can be then stated as follows: Italian and French differ in a systematic way on the pronunciation (*vs*. non-pronunciation) of the definite determiner in (32b), the one which heads degree phrases, either obeying (33) or (34).

(33) In superlatives, only the definiteness of the host DP can and must be overtly expressed, not the one of the DegrP. [Ita]
   a. a max operator over degrees cannot be spelled out as an overt definite determiner.
   b. a covert max operator over degrees is licensed when it's under the scope of another max operator (definite determiner, wh-phrase)

(34) In superlatives, the definiteness of the DegrP must always be overtly expressed. [Fre]
   ≈ null max operators over degrees are never licensed

From (33) and (34), it follows that Italian *la più bella* is only compatible with an attributive superlative with a null noun. On the contrary French 'la plus belle' is only compatible with a predicative LF, where the definite determiner is functioning as a max operator over degrees (as shown in (37b)):

(35) la più bella
   a. $[_{DP}$ la $[_{AP}$ più bella] N]
   b. *$[_{AP}$ $[_{DEGP}$ la più DEGREE] bella]

(36) la plus belle
   a. $[_{DP}$ la $[_{AP}$ plus belle] N]
   b. $[_{AP}$ $[_{DEGP}$ la plus DEGREE] belle]

To conclude, I agree with Kayne 2008 that the same structure can be assigned to both Italian and French postnominal superlatives and that they only differ in the pronunciation vs. non-pronunciation of D. But *contra* Kayne 2008, I argue that the determiner which has to be expressed in French but cannot be expressed in Italian is the one inside the predicate which refers to the unique maximal DEGREE.

(37) a. la fille la plus belle Fre
   b. $[_{DP}$ la … fille … $[_{AP}$ $[_{DegP}$ la plus DEGREE] belle]]

(38) a. la ragazza più bella Ita
   b. $[_{DP}$ la … ragazza … $[_{AP}$ $[_{DegP}$ Ø più DEGREE] bella]]

3.3 Why not a null noun

In the previous subsection, I claimed that postnominal superlatives can be given a truly predicative logical form (*contra* Matushansky 2008). In this section I support this claim against a natural alternative, namely that postnominal superlatives are

attributive superlatives with a null noun (to be understood as identical to the overt noun) as in (39b) (to be compared with my (37b))

(39) a. la fille la plus belle FRE
b. [$_{DP}$ la … fille … [$_{DP}$ la plus belle N]]

I would like to present two pieces of evidence for my analysis: a first one coming for copular alternation in Spanish (which pattern with Italian, obeying (33)) and a second one coming from relative interpretations of what I consider to be predicative superlatives. Let me present the Spanish facts first.

In Spanish, sentential copular sentences involving superlative predicates require the copula *ser*, excluding *estar*. This holds true, regardless of the adjective choice, as shown in (40b).

(40) a. María está enojada
Maria is annoyed.F.SG.
b. *María está la más enojada
Maria is the.F.SG. more annoyed.F.SG.

As reported by Matushansky 2008, this restriction does not extend to DP-internal superlatives. Compare (41) with (40b):

(41) la que está (*la) más enojada
the.F.SG. that is the more annoyed.F.SG.
'the one who is the most annoyed' adapted from Matushansky 2008

I take this to show that the copula *estar*, which has been claimed not to be able to take nominals[8] (see Roy 2006 a.o. for discussion) is only compatible with purely predicative superlatives. Thus, whereas *la más enojada* in (40b) is an attributive superlative containing a null noun, *mas enojada* in (41) is a predicate without nominal projection, as shown here:

(42) a. * …está [$_{DP}$ la más enojada N] (40b)
b. …está [$_{AP}$ [$_{DegP}$ Ø más DEGREE] enojada]] (41)

Copular alternation can only be tested with full relative clause structures, but I assume postnominal superlatives to be reduced relative clauses, minimally different from full structures.

8. Nominals need to be introduced by a preposition in order to be able to appear with *estar*

(i) Juan está *(de) profesor (en USC)
Juan is of professor (at USC)
'Juan is a professor (USC)' adapted from Roy 2006

Moreover, assuming the presence of a null noun is difficult to justify semantically and it would often derive the wrong interpretation. Consider the following French example, which is ambiguous between two interpretations, depending on whether what is compared are years or people.

(43) l' année où Marie fut *(la) plus heureuse
the year where Marie was the more happy
a. the year when M. was happier than any other (relevant) female person
b. the year where Maria was happier than any other year

Now, if a null noun in the LF corresponding to the (a)-reading derives the desired interpretation (where Mary is the happiest PERSON/GIRL), the same doesn't extend to the (b)-reading. Interestingly, in languages obeying (33), the two interpretations are distinguished based on the presence (or absence) of a definite determiner in the predicate:

(44) a. L' anno in cui Maria fu più felice
the year (in) which Maria was more happy
≈ the year where Maria was happier than any other year
b. L' anno in cui Maria fu la più felice
the year (in) which Maria was the.F.SG more happy
≈ the year when M. was happier than any other (relevant) female person

If the presence and the impossibility of an overt determiner in languages like Italian are a reliable way of detecting the presence (or lack thereof) of a nominal projection, then we could argue that in (44a) the superlative has a truly predicative form, without nominal projection. Moreover, if we look at English, the counterpart of (44a) is not compatible with the presence of the noun 'one'. Thus, *the year when Mary was the happiest one* can only have the interpretation in (44b).

3.4 The composition of postnominal superlatives

In Section 2, I claimed that prenominal superlatives are individual-based superlatives, that is individuals are compared and the definite determiner preceding the superlatives plays the canonical role. In the case of postnominal superlatives, I assumed a different derivation, where a relative clause over degree is formed as part of the derivation. This results in a composition where the superlative compares degrees rather than individuals and the (null) definite determiner preceding *plus*/*più* reflects the uniqueness of a degree phrase.

Let me informally show how that works for the postnominal French counterpart of (22) and how this compares to the derivation proposed in 2.4.

(45) le chat blanc le plus lourd
the cat blanc the more heavy
'the heaviest white cat'

First, the comparison class in this case will be the set of different weights of the relevant cats, rather that cats directly. The postnominal superlative then results in a degree relative that refers to the unique maximal weight associated with some cat.[9] When (45) is interpreted in its absolute interpretation, the higher determiner picks out the unique white cat associated with the maximal degree denoted by the predicate.

Note that postnominal superlatives are maximalizing relative clauses of an unusual type. Their maximal interpretation depends on the process of relativization of the subject.

(46) *Une fille la plus belle
a girl the more beautiful

(47) *Una ragazza Ø più bella
a girl the more beautiful

This is something I am not able to offer an explanation for at this point.

As a closing remark, note that unlike prenominal superlatives, postnominal ones can associate with focus and yield relative readings. This is shown in (48) that the reader can compare to (21a).

(48) La sorella più bella, ce l' ha Luigi.
The sister more beautiful, CE CL has Luigi
'LUIGI has the most beautiful sister'

3.5 Summary

In this section, I claimed that postnominal superlatives are predicates in Romance. They are reduced relative clauses with no nominal projection. They are equipped with a MAX operator over degrees which is spelled out as an overt definite determiner in French and is left unpronounced in languages like Italian or Spanish. In this respect, postnominal superlatives pattern together with quantity superlatives and adverbial superlatives.

9. This could be broken down into two steps parallel to what I suggested for prenominal superlatives: (i) the formation of a ordered list of weights and (ii) the selection of the first-ranked element.

4. Conclusion

In this paper, I argued that prenominal superlatives (and arguably elliptical superlatives) are derived through movement of the adjective to a scope position in the left periphery of the DP (to a attributive position in the region of numerals). Postnominal superlatives – on the other hand – are generated as reduced-relative-clause predicates which necessarily extrapose in Romance. They are maximalizing relative clauses over degrees, whose *max* operator has to be overt in French and null in Italian.

References

Cardinaletti, Anna. 1998. "On the Deficient/Strong Opposition in Possessive Systems." In *Possessors, Predicates and Movement in the Determiner Phrase*, ed. by Artemis Alexiadou, and Chris Wilder, 17–53. Amsterdam: John Benjamins. https://doi.org/10.1075/la.22.03car

Cinque, Guglielmo. 2005. "Deriving Greenberg's Universal 20 and its Exceptions." *Linguistic inquiry* 36 (3): 315–332. https://doi.org/10.1162/0024389054396917

Cinque, Guglielmo. 2010. *The Syntax of Adjectives: A Comparative Study*, Vol. 57. MIT Press. https://doi.org/10.7551/mitpress/9780262014168.001.0001

Grasso, Daniele. 2007. "Innovazioni sintattiche in italiano alla luce della nozione di calco." PhD thesis. Université de Genève.

Heim, Irene. 1985. "Notes on Comparatives and Related Matters". Ms., University of Texas, Austin.

Heim, Irene. 1999. "Notes on Superlatives". Ms., MIT.

Kayne, Richard S. 2008. "Some Preliminary Comparative Remarks on French and Italian Definite Articles." *Current Studies in Linguistics Series* 45: 291.

Krasikova, Sveta. 2012. "Definiteness in Superlatives." In *Logic, Language and Meaning*, 411–420. Springer. https://doi.org/10.1007/978-3-642-31482-7_42

Matushansky, Ora. 2008. "On the Attributive Nature of Superlatives." *Syntax* 11 (1): 26–90. https://doi.org/10.1111/j.1467-9612.2008.00101.x

Rohena-Madrazo, Marcos. 2007. *Superlative Movement in Puerto Rican Spanish and General Spanish*. Tech. rep. NYU Working Papers in Linguistics.

Roy, Isabelle. 2006. "Non-verbal Predication: A Syntactic Account of Predicational Copular Sentences." Ph.D. thesis. University of Southern California.

Sharvit, Yael, and Penka Stateva. 2002. "Superlative Expressions, Context, and Focus." *Linguistics and Philosophy* 25 (4): 453–504. https://doi.org/10.1023/A:1020875809794

Sleeman, Petra. 1996. "Licensing Empty Nouns in French." Holland Academic Graphics/HIL Dissertations nr 19.

Szabolcsi, Anna. 1986. "Comparative Superlatives." *MIT Working papers in Linguistics* 8: 245–265.

On recomplementation, high adverbs and V-movement in Spanish

Gabriel Martínez Vera
University of Connecticut

I address the structure of the Spanish left periphery regarding recomplementation, i.e., double-complementizer constructions. I establish a contrast between Latin American Spanish (LAS) and European Spanish (ES): while there is an overt low complementizer *que* in ES, there is a null low complementizer in LAS – this is a novel empirical observation. The latter is different from the construction where low *que* is absent in ES: the null low complementizer in LAS patterns with the overt low complementizer in ES with respect to the phenomena discussed in this paper. I further tie this difference to the height of V-movement and adverb placement in LAS and ES. I then propose a phase-based account of the contrast.

Keywords: complementizer, left periphery, phases, adverbs, Spanish

1. Introduction

In this paper, I address the structure of the Spanish left periphery comparing Latin American Spanish (LAS) and European Spanish (ES). I focus on recomplementation, i.e., double-complementizer constructions. (1) exemplifies this with an overt low complementizer for ES.

(1) Susi dijo que *a los alumnos* **que** les compraron un coche. (ES)
Susi said that the students that CL bought a car
'Susi said that they bought the students a car.'

The main empirical point argued for in this paper is that while there is an overt low complementizer *que* in ES, there is a null low complementizer in LAS – which is the null counterpart of the low complementizer *que* in ES. The construction in LAS is different from the construction where low *que* is absent in ES (in constructions like (1)) in that the null low complementizer in LAS patterns with the overt low complementizer in ES, specifically, with regard to reconstruction effects and the presence of negative constituents. I further link this difference to the height of

https://doi.org/10.1075/rllt.16.12mar

V-movement and adverb placement in LAS and ES.[1] I then propose an account in terms of phases, in particular, I derive this contrast in terms of labeling and antilocality.

The paper is organized as follows: in Section 2, I contrast double-complementizer constructions in ES and LAS; in Section 3, I address the effects of high adverbs in double-complementizer constructions in both ES and LAS; in Section 4, I present my proposal; Section 5 is the conclusion.

2. Contrast between ES and LAS

In this section, I address the basic contrast between ES and LAS recomplementation. I first introduce it in ES and then show how LAS fares regarding the properties the former shows. The examples in (2a,b) illustrate double-complementizer constructions in ES, where a low complementizer *que* can be present or absent (Rodríguez-Ramalle 2003; Villa-García 2012, 2015) – I represent the latter with ø.[2] Between the two complementizers, there is a constituent, a left dislocate (LD), which is a topic in all cases (López 2009b; Villa-García 2012, 2015) – I represent LDs in *italics*.[3] In (2) the LD is *a los alumnos* 'the students'. I refer to the construction in (2a) as the ES *que*-construction and to the one in (2b) as the ES ø-construction:[4]

(2) a. Susi dijo que *a los alumnos*, **que** les dieron regalos. (ES)
b. Susi dijo que *a los alumnos* **ø** les dieron regalos. (ES)
Susi said that the students **that** CL gave presents
'Susi said that they gave the students presents.'

As noted by Villa-García (2012, 2015), (2a,b) differ minimally regarding reconstruction effects, specifically, bound variable (3), anaphor binding (4) and scope reconstruction (5). While these effects are found in the ES ø-construction,

1. My data survey of LAS includes speakers from Mexico, Puerto Rico, Colombia, Ecuador, Peru and Argentina. For ES, I am using data reported in the literature, but have also gathered data from speakers of Asturias, Madrid, Barcelona and the Basque Country.

2. Villa-García (2012, 2015) builds on an extensive literature on recomplementation in ES. Recent discussion appears in Demonte and Fernández-Soriano (2007 et seq.), López (2009a), Rodríguez-Ramalle (2003), among many others. I refer to Villa-García's work for the most part for simplicity, as he summarizes the previous literature to a large extent.

3. More precisely, all the cases on which I focus are instances of clitic left dislocation (CLLD) (López 2009b; Villa-García 2012; 2015).

4. Here I focus on embedded clauses. See López (2009b) for discussion of matrix clauses.

they are absent in the ES *que*-construction: in (3), the LD *a su perro* 'his dog' cannot be bound in the ES *que*-construction (3a), but it can be in the ES ø-construction (3b).

(3) a. Yo dije que *a su perro*$_{*i/j}$, **que** todo el mundo$_i$ lo pasea. (ES)
b. Yo dije que *a su perro*$_{i/j}$ **ø** todo el mundo$_i$ lo pasea. (ES)
I said that his dog **that** all the world CL walks
'I said that everybody walks his dog.'

In (4), the anaphor *a sí misma* 'herself' cannot be bound by *María* in the ES *que*-construction (4a) – yielding a Condition A violation –, but it can be in the ES ø-construction (4b).

(4) a. *Yo dije que *a sí misma*$_i$, **que** María$_i$ se envía correos a diario. (ES)
b. Yo dije que *a sí misma*$_i$ **ø** María$_i$ se envía correos a diario. (ES)
I said that herself **that** Mary CL sends emails every day
'I said that Mary sends herself emails every day.'

In (5) below, the question is whether the universal (*todos los edificios* 'all the buildings') can take scope over the numeral (*dos chicos* 'two boys') – the narrow scope reading is available in both cases.[5] While this is impossible in the ES *que*-construction (5a), it is possible in ES ø-construction (5b). This suggests that, while the LDs *a su perro* 'his dog', *a sí misma* 'herself' and *dos chicos* 'two boys' are base-generated in their surface position in (2a,3a,4a) respectively, they move to this position in (3b,4b,5b), so reconstruction to a position lower to that of the antecedent becomes possible (López 2009b; Villa-García 2012, 2015).[6,7]

(5) a. Dije que *a dos chicos*, **que** los sacaron de todos los edificios. (ES)
b. Dije que *a dos chicos* **ø** los sacaron de todos los edificios. (ES)
said that two boys **that** CL rescued from all the buildings
'I said that two boys were rescued from every building.'

5. For the narrow scope case, in which the numeral takes scope over the universal, imagine a context in which the same two boys are rescued from different buildings in several days.

6. This has been challenged by, e.g., Boeckx (2007). See Villa-García (2015) for discussion.

7. See Villa-García (2015) for discussion of additional reconstruction paradigms.

In addition to these contrasts, the facts concerning negation also suggest that the LD is base-generated in the left periphery in the ES *que*-construction, but not in the ES ø-construction. I assume an analysis where NegP is located below the projection headed by low *que* (Villa-García 2015; see also López 1995; Zanuttini 1997). I further assume, following Bošković (2001), that Neg is a null affix such that a negative constituent and this null affix have to be adjacent at some point in the derivation.[8] If a constituent of this sort is not available, the negative element *no* is inserted – hence, this is just like the English *do* support. Thus, in (6a) *nadie* 'nobody' is adjacent to Neg (at some point of the derivation), so it hosts the affix and *no* is not inserted. In (6b), however, *no* is inserted due to the lack of a host – the negative phrase is never adjacent to Neg in the derivation. The contrast in (6) holds in ES and LAS.

(6) a. Nadie (*no) vino.
b. *(No) vino nadie. (ES/LAS)
(no) came nobody
'Nobody came'

In recomplementation, the prediction is that *no* will appear in a sentence only if an LD with a negative constituent is not adjacent to Neg at any point in the derivation. In the ES *que*-construction (7a), *no* is present, suggesting that the LD *a ninguno de los chicos* 'none of the children' is base-generated in the left periphery. In contrast, in the ES ø-construction (7b) *no* appears, which suggests that the LD moves to the left periphery from a lower position – it is then adjacent to Neg at one point of the derivation (Villa-García 2015).

(7) a. Dijo que *a ninguno de los chicos*, **que** Juan *(no) los invitó (ES)
b. Dijo que *a ninguno de los chicos* **ø** Juan (*no) los invitó. (ES)
said that none of the children **that** John no CL invited
'S/he said that John didn't invite any of the children.'

8. An implementation of this idea is in terms of agreement between the Spec and the head. See Haegeman & Zanuttini (1996) (see also Zeiljstra 2007).

In LAS, (2a), repeated in (8a), is ungrammatical – there cannot be an overt low *que* in LAS. Examples that on the surface look like (2b), with a non-overt low *que*, are possible (8b) – I use C_{null} (and, correspondingly, LAS C_{null}-construction) for the former and keep **ø** for ES.[9]

(8) a. *Susi dijo que *a los alumnos*, **que** les dieron regalos. (LAS)

b. Susi dijo que *a los alumnos*, $\mathbf{C_{null}}$ les dieron regalos. (LAS)
Susi said that the students **that** CL gave presents
'Susi said that they gave the students presents.'

Crucially, LAS (9), with only one overt *que*, patterns with (3a,4a,5a), not (3b,4b,5b), regarding bound variable reconstruction (9a), anaphor binding (9b) and scope reconstruction (9c), i.e., these effects are not possible: in (9a) *a su perro* 'his dog' cannot be bound by *todo el mundo* 'everybody'; in (9b), *a sí misma* 'herself' cannot be bound by *María*; and in (9c), only the narrow scope reading, where *dos chicos* 'two boys' takes scope over *todos los edificios* 'all the buildings', is possible. With negation, LAS patterns with (7a), not (7b), as in (9d): the negative word *no* is inserted (thus, the LD with a negative element is never adjacent to Neg in the derivation).

(9) a. Yo dije que *a su perro*$_{*i/j}$, $\mathbf{C_{null}}$ todo el mundo$_i$ lo pasea (LAS)
I said that his dog **that** all the world CL walks
'I said that everybody walks his (=somebody else's) dog.'

b. *Yo dije que *a sí misma*$_i$, $\mathbf{C_{null}}$ María$_i$ se envía
I said that herself **that** Mary CL sends
correos a diario. (LAS)
emails every day
'I said that Mary sends herself emails every day.'

9. In the examples, an intonational break is included with a comma, following Villa-García's (2015) convention (see also, among many others, Rizzi 1997). For LAS, I only focus on constructions with such break. Due to space considerations, I will not discuss an alternative construction where the break is absent, which shows different properties. There are also some cases in which a low *que* may surface in LAS, e.g., with *wh*-words like *cuándo* 'when' in the sequence LD + *que* + *wh*-word. Such constructions lie beyond the scope of this paper.

c. Dije que *a dos chicos*, $\mathbf{C}_{\text{null}}$ los sacaron de todos
said that two boys **that** CL rescued from all
los edificios. (LAS)
the buildings
'I said that two boys were rescued from every building.'

d. Dijo que *a ninguno de los chicos*, $\mathbf{C}_{\text{null}}$ Juan *(no)
said that none of the children **that** John no
los invitó. (LAS)
CL invited
'S/he said that John didn't invite any of the children.'

To sum up, (i) LAS has a rich left periphery when recomplementation is considered, just as ES; (ii) LAS (8), with C_{null}, should be analyzed like ES (1a), in which low *que* is present; and (iii) missing secondary *que* constructions in LAS (C_{null}) and ES (ø) should be treated differently.

3. High adverbs and height of V-movement

In this section, I address high adverbs in double-complementizer constructions; their distribution will be key to the analysis in Section 4. The literature for ES recomplementation (Villa-García 2015) shows that LDs are not restricted to arguments (in all the examples in Section 2, the LDs are arguments), but can include, for instance, adjuncts, including various kinds of adverbs (see also Benincà 1988; Rizzi 1997).[10] In addition, there can be more than one LD in double-complementizer constructions (not only one, as in the examples in Section 2).

The possibility of there being more than one LD is shown in (10,11). Here the LDs are an argument and a high adverb (Cinque 1999, 2004). The argument is *al chico* 'the boy'; the high adverb used is *afortunadamente* 'fortunately' (similar adverbs, e.g., *felizmente* 'happily' also work). The order of these two is interchangeable; in particular, the adverb LD can appear following (10) and preceding (11) the argument LD. As can be seen in the (c) examples, these properties also hold in LAS.

10. The range of LDs is greater than discussed in this paper. See Villa-García (2015).

(10) a. Dijo que *al chico*, *afortunadamente*, **que** le dieron regalos. (ES)
b. Dijo que *al chico*, *afortunadamente* **ø** le dieron regalos. (ES)
c. Dijo que *al chico*, *afortunadamente*, $\mathbf{C}_{\text{null}}$ le dieron regalos. (LAS)
said that the boy fortunately **that** CL gave presents
'S/he said that fortunately they gave the boy presents.'

(11) a. Dijo que *afortunadamente*, *al chico*, **que** le dieron regalos. (ES)
b. Dijo que *afortunadamente*, *al chico* **ø** le dieron regalos. (ES)
c. Dijo que *afortunadamente*, *al chico*, $\mathbf{C}_{\text{null}}$ le dieron regalos (LAS)
said that fortunately the boy **that** CL gave presents
'S/he said that fortunately they gave the boy presents.'

Of relevance here is that *afortunadamente* 'fortunately' cannot follow the verb in the ES *que*-construction (12a). This holds for the LAS C_{null}-construction (12b) as well. Crucially, it can follow the verb in the ES ø-construction (13).

(12) a. ?*Dijo que *al chico*, **que** le dieron, *afortunadamente*, regalos. (ES)
b. ?*Dijo que *al chico*, $\mathbf{C}_{\text{null}}$ le dieron, *afortunadamente*, regalos. (LAS)
said that the boy **that** CL gave fortunately presents
'S/he said that fortunately they gave the boy presents.'

(13) (?)Dijo que *al chico*, le dieron, *afortunadamente*, regalos. (ES)
said that th boy CL gave fortunately presents
'S/he said that fortunately they gave the boy presents.'

The contrast in (12, 13) suggests that the ES *que*-construction, and the LAS C_{null}-construction pattern alike – and contrast with the ES ø-construction. In particular, the contrast suggests that the landing site for V-movement is different. In the ES *que*-construction and in the C_{null}-construction, the verb cannot move past the adverb (12), whereas in the ES ø-construction it can (13). This suggests that the verb moves higher in the ES ø-construction than in the ES *que*-construction and in the LAS C_{null}-construction. Note that this discussion confirms the split discussed in section 2 regarding reconstruction effects. The ES *que*-construction and the C_{null}-construction pattern together with respect to adverb placement and reconstruction effects: the verb cannot precede the adverb and reconstruction

effects are not possible. The ES ø-construction behaves differently: the verb can precede the adverb and reconstruction effects are possible. Note also that the contrasts discussed do not lie in the overt vs. null nature of the complementizers, as overt low *que* in ES and C_{null} in LAS pattern together, and ø in ES patterns differently.

4. Analysis

In this section, I propose an analysis that accounts for the facts discussed in Sections 2 and 3. The structure I assume is based on Rizzi (2001), as applied to ES by Rodríguez-Ramalle (2003) and Villa-García (2012, 2015) (see also Paoli 2007), where the low C heads TopicP – *que*/ø in ES and C_{null} in LAS (FinP=FinitenessP):

(14) $[_{ForceP}\ [_{Force'}$ que $[_{TopicP}$ LD $[_{Topic'}$ **que / ø / C_{null}** $[_{FinP}\ [_{Fin'}$ … V …]]]]]]

The gist of the proposal is that, in the dislocation structure with ø in ES, the verb moves to the position where low *que* is located when present in ES. This voids the phasehood of the projection below *que* (Gallego & Uriagereka 2007; den Dikken 2007; Bošković 2015). This V-movement (and phase voiding) does not take place in the ES *que*-construction. The verb does not move to the position where overt low *que* is located in ES, so the phasehood of the projection below *que* is not voided. Importantly, LAS behaves like the ES *que*-construction, not like the ES ø-construction in that the verb does not move to the position where C_{null} is located – despite the similarity on the surface between the ES ø-construction and the LAS C_{null}-construction in that both complementizers are null.

Voidance of phasehood takes place via head movement, as argued for in Gallego & Uriagereka (2007), den Dikken (2007) and Bošković (2015): head movement of head X of phase XP voids phasehood of XP.[11] In recomplementation, assuming the structure in (14), the proposal is that the verb never moves to the Topic head in LAS (8b)/(15c) and fails to move there with overt low *que* in ES (2a)/(15a). The

11. These proposals differ in that for Gallego & Uriagereka (2007) and den Dikken (2007) the higher phrase becomes a phase, whereas for Bošković (2015) it does not. For my account, the phasehood status of the higher phrase does not matter. What is relevant is that the lower phase is gone. For this reason, I adopt neutral terms: I refer to this as *voidance of phasehood*.

verb does move to this position when there is no overt low *que* in ES – i.e., the ES ø-construction in (2b)/(15b).[12]

(15) a. Susi dijo que *a los alumnos*, **que** les dieron regalos. (ES)
b. Susi dijo que *a los alumnos* **ø** les dieron regalos. (ES)
c. Susi dijo que *a los alumnos*, $\mathbf{C}_{\text{null}}$ les dieron Regalos (LAS)
Susi said that the students **that** CL gave presents
'Susi said that they gave the students presents.'

Furthermore, movement past low *que* in ES and C_{null} in LAS, essentially yields a *Comp*-trace effect (this is not the case with ø in ES, where there is no movement to Spec,FinP) – I come back to this below in this section. As has been claimed in the literature (see Holmberg 2001; López 2009b; contra Bošković 2015), I assume that FinP is a phase, which forces successive cyclic movement to the edge of FinP.[13] (16) illustrates the contrast: in (16a), with low *que* for ES and C_{null} for LAS, movement of the LD from Spec,FinP is banned, while, in (16b), with ø for ES, movement of the LD from a lower position (crucially, not Spec,FinP, since the phasehood of FinP is voided) is possible – I use standard X-bar notation (which will be revised below); the arrows only illustrate the relevant movement of the LD:[14]

---

12. Note that my proposal claims that the verb moves to a low position in the left periphery in one particular way (see the structures in (16) below). As early as Suñer (1994), contra Torrego (1984), it was claimed that the verb does not move to C (in questions), but stays in I – crucially, relying on the position of adverbs. An interesting venue for research on this matter would be not to ban (or allow) movement to C in an all-or-nothing fashion, but to qualify it in terms of which layer in the split CP (and when) can be targeted by V-movement.

13. Assuming that phases can be elided (see López 2009 and Villa-García 2015 for discussion in Romance), the fact that FinP can be elided suggests that it is a phase (C=*que*/ø/C_{null}).

(i) Dije que a María, **C** [la llamaron] y que a José, **C** [~~la llamaron~~]
said that Mary **that** CL called and that Joseph **that** CL called
también.
also
'I said that they called Mary and also Joseph.'

14. Note that the height of V-movement in matrix imperatives, e.g., in which clitics are involved (in particular, with the Verb-Clitic order), does not affect my proposal, since V-movement in imperatives may involve a higher position in the left periphery. Note also that there

(16) a. … [$_{\text{TopicP}}$ LD [$_{\text{Topic'}}$ **que / C$_{\text{null}}$** [$_{\text{FinP}}$ ~~LD~~ [$_{\text{Fin'}}$ … V … ~~LD~~ …]]]]
*

b. … [$_{\text{TopicP}}$ LD [$_{\text{Topic'}}$ **V$_i$** [$_{\text{FinP}}$ [$_{\text{Fin'}}$ … **~~t$_i$~~** … ~~LD~~ …]]]]
ok

The question now is: how is the locality effect under discussion derived? For this, I apply Bošković's (2016) account of the *Comp*-trace, which deduces it from antilocality (see Erlewine 2016 for a similar approach), in conjunction with the labeling algorithm (Chomsky 2013), which is defined in (17):

(17) *Labeling Algorithm*
a. If the head H and a non-head XP are merged to form the syntactic object SO={H, XP}, H labels SO.
b. If two non-heads XP and YP are merged to form the syntactic object SO={XP, YP}, there are two cases to consider:
i. If XP and YP share a feature, such feature provides the label.
ii. If XP and YP do not share a feature, the syntactic object SO created is not labeled; when XP or YP moves, the remaining phrase labels SO – assuming traces are ignored for labeling.

I illustrate (17) with the examples in (18).

(18) a. [$_{\text{VP}}$ arrived the man]
b. I wonder [$_{\text{QP}}$ [who$_{i\,[+Q]}$] [$_{\text{CP}}$ C$_{[+Q]}$ [$_{\text{IP}}$ she kissed t$_i$]]
c. Who$_i$ do you think [$_{\text{CP}}$ t$_i$ that she kissed t$_i$]
c′. …. [$_{?}$ [$_{\text{QP}}$ who$_{i\,[+Q]}$] [$_{\text{CP}}$ that … t$_i$]

(18a) exemplifies (17a): there is a head, the verb *arrived*, and a DP, *the man*, so the head labels the new SO as VP. (18b) exemplifies (17bi): the complementizer (i.e., CP) and the *wh*-phrase share a feature, i.e., [+Q], which labels the new SO as QP. (18c,c′) illustrates (17bii): (18c′) shows the intermediate step in which two phrases are combined ({QP, CP}); no feature is shared here ([+Q] is present in the *wh*-word, but not in C), so the new SO cannot be labeled – I represent unlabeled projections with a question mark (?). The SO can only be labeled when one of the phrases moves; in this case, the *wh*-phrase moves. The unlabeled SO can then be labeled as CP (18c). Note that the case in (18c) holds under the assumption that successive cyclic movement does not involve agreement (see, e.g., Chomsky 2013),

are accounts where the Verb-Clitic order in imperatives has nothing to do with V-movement (see, for instance, Bošković 2004, who argues that such order arises due to a PF effect, i.e., pronunciation of lower copies of clitics is forced by morphological merger between the verb and the imperative affix).

which means that the relevant element, i.e., SO={QP, CP}, cannot be labeled – no agreement is involved in the cases I consider.

The second element needed to derive the *Comp*-trace effect, and the effect discussed here, is antilocality – Bošković (2016) defines it in the labeling framework. (19) says that moved elements must cross a labeled projection:

(19) *Antilocality*
Movement of α targeting β must cross a projection distinct from β
(unlabeled projections do not count as distinct from labeled ones).

To show how (17,19) work, consider the abstract structure in (20) involving successive cyclic movement:

(20) [YP [$_{ZP}$ Z [$_{?}$ YP [$_{XP}$ X [$_{WP}$ …]]]]]
*

Here, the heads X and Z provide labels (XP and ZP), as they appear in a SO={H, XP} configuration – this is the case of (17a). In contrast, merging YP and XP creates an unlabeled object – this falls under the case in (17bii), where no feature sharing is involved. Movement of YP targeting ZP is ruled out by (19), since YP does not cross a labeled projection. As discussed in Bošković (2016), (21) represents the English *Comp*-trace effect, e.g., subject extraction from embedded clauses in questions, where ZP=CP, Z=*that*, YP=*wh*-word and XP=IP, as shown below. There is an antilocality violation, as the *wh*-word has to move to the edge of the phase, CP in this case, but such movement does not cross a labeled projection, violating antilocality:[15]

(21) … [*wh*-word [$_{CP}$ that [$_{?}$ *wh*-word [$_{IP}$ I …]]]]
*

As is well known, however, Spanish does not have a 'standard' *Comp*-trace effect. Nonetheless, although Spanish otherwise does not have a *Comp*-trace effect, recomplementation constructions behave like the English *Comp*-trace effect; the suggestion made here is that the two (i.e., English *Comp*-trace effect and Spanish recomplementation) should be accounted for in a unified way (see also Villa-García 2015). The 'standard' *Comp*-trace effect is tied to the obligatory movement of the subject to Spec,IP in English – e.g., to satisfy an EPP feature. Subjects

15. This is also the reason why there is no short subject topicalization in English, as movement from Spec,IP to CP does not cross a labeled projection, violating antilocality (see Bošković 2016; see also this work regarding labeling within the IP projection, which crucially takes place after the subject moves from Spec,IP, hence the ? in (21) at the point of movement).

in Spanish are not required to move to this position, thus voiding the 'standard' *Comp*-trace effects.

Crucially, in recomplementation, movement to the edge of FinP is required because FinP is a phase, creating the same configuration English shows to create a *Comp*-trace effect. There are two cases two consider here, i.e., (i) low *que* in ES and C_{null} in LAS, and (ii) ø in ES. With low *que* in ES and C_{null} in LAS, the LD has to move to the edge of FinP (because it is a phase), so the new projection is unlabeled (22). The movement to Spec,TopicP yields an antilocality violation, as no projection different from TopicP is crossed – this is precisely an instantiation of the structure in (20).

(22) [LD [$_{TopicP}$ **que / C$_{null}$** [$_{?}$ ~~LD~~ [$_{FinP}$ … ~~LD~~]]]]
*

In the case of ø in ES, on the other hand, the LD can move to Spec,TopicP directly, since V-movement from FinP to TopicP voids the phasehood of the former, so the LD need not stop at the edge of FinP in (23):

(23) [LD [$_{TopicP}$ **V$_i$** [$_{FinP}$ … ~~t$_i$~~ … LD …]]]
ok

As anticipated in section 3, these differences cannot lie in the overt or null nature of the C (as suggested by Villa-García 2012, 2015 for ES alone) due to the contrast in the ES ø-construction and the LAS C_{null} construction.

The account derives the possibility/impossibility of reconstruction effects in (3–5) for ES and in (9a–c) for LAS, as well as the contrast in the presence/absence of negation in ES (7) and LAS (9d). These are synthetized in the table in (24):

(24) *Contrasts in ES and LAS recomplementation*

	ES	**LAS**	**ES**
	que	**C$_{null}$**	**ø**
Bound variable	No	No	Yes
Anaphor binding	No	No	Yes
Scope reconstruction	No	No	Yes
Presence of negation	Yes	Yes	No

The schematic account in (22) is the relevant structure here. Since movement from the edge of FinP to the edge of TopicP yields an antilocality violation, the LD cannot be generated in a lower position; it can only be base-generated in its surface

position. Thus, no reconstruction is possible – no position where the LD can be reconstructed is available. I illustrate this with bound variable reconstruction. (25a,b) rewrite (3a,9a); (25c) shows that the movement derivation is not possible.

(25) a. Yo dije que *a su perro*$_{*i/j}$, **que** todo el mundo$_i$ lo pasea. (ES)

b. Yo dije que *a su perro*$_{*i/j}$, **C$_{null}$** todo el mundo$_i$ lo pasea. (LAS)
I said that his dog **that** all the world CL walks
'I said that everybody walks his (=somebody else's) dog.'

c. [his dog [$_{TopicP}$ **que / C$_{null}$** [$_?$ ~~his dog~~ [$_{FinP}$ … ~~his dog~~]]]]
*

In contrast, in the ES ø-construction, reconstruction is possible (cf. (3b,4b,5b)), as the verb moves to Topic voiding the phasehood of FinP. The LD can thus be base-generated in a lower position and move to Spec,TopicP without stopping at the edge of FinP. There is then a lower position that can be reconstructed into, so bound variable, anaphor binding and scope reconstruction effects are possible. I show this with bound variable reconstruction. (26a) repeats (3b); (26b) illustrates the movement derivation.

(26) a. Yo dije que *a su perro*$_{i/j}$, **ø** todo el mundo$_i$ lo pasea.(ES)
I said that his dog **that** all the world CL walks
'I said that everybody walks his dog.'

b. [his dog [$_{TopicP}$ **walk$_i$** [$_{FinP}$ … ~~**t$_i$**~~ … ~~his dog~~ …]]]
ok

The analysis also captures the presence of the negative element *no* in the ES *que*-construction in (7a) and the LAS C$_{null}$-construction in (9d). Since the LD with a negative element is base-generated in the left periphery, it is never adjacent to Neg, so the negative element *no* surfaces. In contrast, in (7b) in the ES ø-construction, the LD with a negative element moves to the left periphery from a lower position, being adjacent to Neg at some point in the derivation. Thus, in this case, *no* does not surface.

Based on this discussion, it is in fact possible that ø (ES) and C$_{null}$ (LAS) are the same head. The difference would lie in the possibility of V-movement to the head in question in ES. The difference would thus reduce to the height of V-movement in each construction.[16] I further suggest that it is also possible that the head of TopicP must be lexicalized in ES – as discussed in the two options in (22,23):

16. I leave discussing the motivation for this difference for future research.

this could be done either via the presence of low *que* (22) or via V-movement (23) (see Paoli 2007 for a similar suggestion). The lexicalization requirement would not apply in LAS.[17]

5. Conclusion

I have discussed novel recomplementation data addressing variation in two macro-varieties of Spanish. Although only ES has the double *que* construction with an overt low *que*, LAS does show recomplementation. I have provided evidence for this. A null C, C_{null}, in LAS patterns with overt low *que*-constructions in ES (and differently from constructions in ES where low *que* does not appear) regarding reconstruction – bound variable, anaphor binding and scope reconstruction effects – , which is not possible; regarding the presence of *no* when the left dislocate is a negative constituent; and regarding adverb placement – high adverbs cannot follow the verb. I have interpreted this as reflecting a difference in the height of the verb. Low *que* in ES and C_{null} in LAS thus pattern together and differently from ø in ES regarding reconstruction effects and V-movement. I have proposed a uniform phase-based account of these differences. Under the assumption that Fin(iteness)P – the complement of the Topic head whose Spec LD occupies – is a phase, movement to Spec,TopicP must go through Spec,FinP, violating antilocality – as such movement does not cross a labeled projection. This rules out the possibility of a movement derivation in the ES *que*-construction and the LAS C_{null}-construction, which accounts for the lack of reconstruction effects, the presence of the negative element *no* and the fact that the verb cannot precede high adverbs. In the ES ø-construction, in contrast, V-movement to Topic voids the phasehood of FinP. Thus, movement to Spec,TopicP need not proceed via the edge of FinP – thus antilocality is not violated. The availability of the movement derivation accounts for the possibility of reconstruction effects, the abscence of the negative element *no* and the fact that the verb can precede high adverbs.

17. An anonymous reviewer points out that the antilocality violation in the structures under discussion (in particular in the ES *que*-construction and in the LAS C_{null}-construction) should not arise if there is an additional projection between TopicP and FinP. One possible instance of this is when there are multiple topics (as the reviewer mentions). Suppose that, in this case, there are multiple TopicPs, as proposed by Villa-García (2012, 2015). The LDs need to move to the edge of FinP before moving to the Spec of the relevant TopicP. Under this approach, while higher LDs can reach the relevant Spec, movement of the lowest one will always yield an antilocality violation (as it crosses FinP, which is a phase).

Acknowledgement

I would like to especially thank Željko Bošković for his constant feedback and suggestions to improve this work. I also thank an anonymous reviewer for helpful comments. All errors remain my own.

References

Benincà, Paola. 1988. "L'ordine degli elementi della frase e le costruzioni marcate." In *Grande Grammatica Italiana di Consultazione*, Vol. 1, ed. by Lorenzo Renzi, 129–194. Bologna: Il Mulino.

Bošković, Željko. 2001. *On the Nature of the Syntax-Phonology Interface*. Amsterdam: Elsevier.

Bošković, Željko. 2004. "Topicalization, Focalization, Lexical Insertion, and Scrambling." *Linguistic Inquiry* 35: 613–638. https://doi.org/10.1162/0024389042350514

Bošković, Željko. 2015. "From the Complex NP Constraint to Everything." *The Linguistic Review* 32: 603–669. https://doi.org/10.1515/tlr-2015-0006

Bošković, Željko. 2016. "On the Timing of Labeling." *The Linguistic Review* 33: 17–66.

Boeckx, Cedric. 2007. *Understanding Minimalist Syntax*. Oxford: Wiley-Blackwell.

Chomsky, Noam. 2013. "Problems of Projection." *Lingua* 130: 33–49. https://doi.org/10.1016/j.lingua.2012.12.003

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: Oxford University Press.

Cinque, Guglielmo. 2004. "Issues in Adverbial Syntax." *Lingua* 114: 683–710. https://doi.org/10.1016/S0024-3841(03)00048-2

Demonte, Violeta, and Olga Fernández-Soriano. 2007. "La periferia izquierda oracional y los complementantes del español." In *Vernetzungen: Kognition, Bedeutung, (kontrastive) Pragmatik*, ed. by Juan Cuartero, and Martine Emsel, Frankfurt: Peter Lang.

den Dikken, Marcel. 2007. "Phase Extension." *Theoretical Linguistics* 33: 1–41. https://doi.org/10.1515/TL.2007.001

Erlewine, Michael. 2016. "Anti-Locality and Optimality in Kaqchikel Agent Focus." *Natural Language & Linguistic Theory* 34: 429–479. https://doi.org/10.1007/s11049-015-9310-z

Gallego, Ángel, and Juan. Uriagereka. 2007. "Sub-Extraction from Subjects." In *Romance Linguistics 2006*, ed. by José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez, and Marí José Cabrera, 149–162. Amsterdam: John Benjamins. https://doi.org/10.1075/cilt.287.12gal

Haegeman, Liliane, and Zanuttini Raffaella. 1996. "Negative Concord in West Flemish." In *Parameters and Functional Heads: Essays in Comparative Syntax*, ed. by Adriana Belletti, and Luigi Rizzi, 117–179. Oxford: Oxford University Press.

Holmberg, Anders. 2001. "The Syntax of Yes and No in Finnish." *Studia Linguistica* 55: 140–174. https://doi.org/10.1111/1467-9582.00077

López, Luis. 1995. "The Hierarchy of Negation and Tense." In *Proceedings of the Eastern States Conference on Linguistics* X: 186–197.

López, Luis. 2009a. "Ranking the Linear Correspondence Axiom." *Linguistic Inquiry* 40: 239–276. https://doi.org/10.1162/ling.2009.40.2.239

López, Luis. 2009b. *A Derivational Syntax for Information Structure*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199557400.001.0001

Paoli, Sandra. 2007. "The Fine Structure of the Left Periphery: COMPs and Subjects: Evidence from Romance." *Lingua* 117: 1057–1079. https://doi.org/10.1016/j.lingua.2006.05.007

Rizzi, Luigi. 1997. "The Fine Structure of the Left Periphery." In *Elements of Grammar: A Handbook of Generative Syntax*, ed. by Liliane Haegeman, 281–337. Dordrecht: Kluwer.

Rizzi, Luigi. 2001. "On the position "Int(errogative)" in the Left Periphery of the Clause." In *Current Studies in Italian Syntax*, ed. by Guglielmo Cinque, and Giampaolo Salvi, 287–296. Amsterdam: Elsevier.

Rodríguez-Ramalle, Teresa. 2003. *La gramática de los adverbios en* -mente *o cómo expresar maneras, opiniones y actitudes a través de la lengua*. Madrid: Ediciones de la Universidad Autónoma de Madrid.

Suñer, Margarita, 1994. "V-Movement and the Licensing of Argumental Wh-Phrases in Spanish." *Natural Language and Linguistic Theory* 12: 335–372. https://doi.org/10.1007/BF00993148

Torrego, Esther. 1984. "On Inversion in Spanish and Some of its Effects." *Linguistic Inquiry* 15: 103–129.

Villa-García, Julio. 2012. "Recomplementation and Locality of Movement in Spanish." *Probus* 24: 257–314. https://doi.org/10.1515/probus-2012-0011

Villa-García, Julio. 2015. *The Syntax of Multiple-que Sentences in Spanish*. Amsterdam/Philadelphia: John Benjamins.

Zanuttini, Raffaella. 1997. *Negation and Clausal Structure: A Comparative Study of Romance Languages*. Oxford: Oxford University Press.

Zeiljsta, Hedde. 2007. "Negation in Natural Language: On the Form and Meaning of Negative Elements." *Language and Linguistics Compass* 1: 498–518. https://doi.org/10.1111/j.1749-818X.2007.00027.x

Articles in an article-less language

Romanian-Serbian code-switching

Vanessa Petroj
University of Connecticut

I will explore the interaction between Romanian and Serbian elements within the Traditional Noun Phrase (TNP) that participate in code-switching, including Romanian definite articles (D), Romanian and Serbian adjectives (A), and Serbian nouns (N), focusing on the definite article affixation. Following Bošković (2008, 2012), I assume Serbian to be an NP language (lacking definite articles) and Romanian a DP language (having definite articles), leading to semantic and syntactic consequences for the rest of the structure. Finally, I argue against CS-specific constraints such as Free Morpheme Constraint (Poplack 1980), Functional Head Constraint (Belazi, Rubin, & Toribio 1994), and PF Disjunction Theorem (MacSwan 1999), and in support of Bandi-Rao and den Dikken (2014) reformulated PF Disjunction Theorem.

Keywords: code-switching, syntax, Romanian, Serbian, DP/NP parameter setting, Free Morpheme Constraint, Functional Head Constraint, PF Disjunction Theorem

1. Introduction

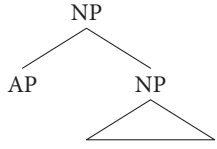
A fundamental assumption for this paper is that there is a structural difference between languages with and without definite articles, as argued by Bošković (2008, 2012, 2014) and other authors (Fukui 1988, Corver 1992, Chierchia 1998, Baker 2003, i.a.).[1] As such, languages without articles lack the DP projection where definite articles are presumed to be positioned, leading to a structural difference that

1. The reviewer has rightfully noted that the views regarding the DP/NP treatment of languages is controversial, with authors like Abney (1987), Bowers (1991), Longobardi (1994), Pereltsvaig (2007), i.a., claiming that DP is universal across all languages. Due to space limitations, I will not further discuss this here, however, I refer the reader to Petroj (2019) and Petroj (to appear) for additional discussion and evidence regarding the different NP status of Romanian and Serbian through CS.

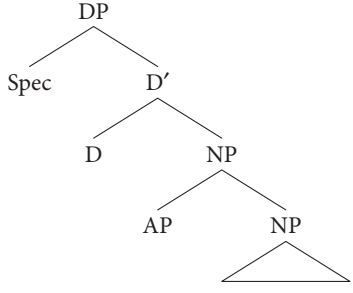
https://doi.org/10.1075/rllt.16.13pet

divides languages into NP (articleless) and DP (languages with articles) languages (Bošković 2008, 2014, i.a.). Thus, TNP is used to refer to the traditional NP that includes crosslinguistic variation with respect to the DP/NP parameter. Therefore, the DP is the TNP in Romanian, and the NP in Serbian. Following Bošković (2008, 2012), I assume Serbian to be an NP language (lacking definite articles, therefore, the DP layer) and Romanian a DP language (having definite articles, and therefore, the DP layer). This is represented structurally in (1a) for Serbian (1b) for Romanian (the structures below take into consideration determiners, adjectives, and nouns):

(1) a.



b.



Due to elements originating in two structurally different languages, when parts of the structures mix during CS, the NP/DP distinction ought to have consequences on the resulting derivation. While the details and implications of CS involving NP/DP languages are not directly relevant here (for a more comprehensive discussion, see Petroj (to appear)), it is important to note that, contrary to claims by Poplack (1980), Belazi, Rubin, and Toribio (1994), and MacSwan (1999), and following to Bandi-Rao and den Dikken's (2014), word-internal CS is quite productive in this language pair.

This paper is outlined as follows. Section 1 introduces fundamental assumptions required for this paper. Section 2 and Section 3 illustrate relevant elements and their distribution within the Romanian DP and Serbian NP respectively. An overview of the CS TNP is laid out in Section 4, while Section 5 offers a novel account for the derivation of relevant CS constructions. Next, Section 6 offers a brief discussion of relevant CS constraints. Finally, Section 7 concludes this chapter. For ease of exposition, I will follow the common practice of marking elements from the two languages uniformly throughout the paper; in CS examples, **Romanian** elements will be in **bold**, and *Serbian* in *italics*.

2. Romanian DP

While Romanian has both indefinite and definite articles, this paper will only focus on contexts where definite articles are used. The Romanian definite article (**-ul**)[2] agrees in number and gender with the noun, and, as a bound morpheme, it requires a phonological host to be realized. This is roughly the linearly closest element in the DP, usually the noun (2a) or adjective (2b), but never both categories within the same DP (3). However, not all elements can act as hosts (e.g. adverbs). This is illustrated in (4), where the article is not hosted by the linearly closest element (i.e. adverb), but by the closest potential host (i.e. adjective).

(2) a. băiat-(u)l frumos
boy.the beautiful

b. frumos-(u)l băiat
beautiful.the boy

(3) a. *băiat-(u)l frumos-(u)l
boy.the beautiful.the

b. *frumos-(u)l băiat-(u)l
beautiful.the boy.the

(4) a. *foarte-(u)l înalt student
very.the tall student

b. foarte înalt-ul student (Rom)
very tall.the student

Like articles, adjectives also agree in number and gender with the noun they modify (5). As seen above, Romanian adjectives can also act as hosts for the definite article, and they can occur on either side of the noun, as in (6), with certain semantic differences[3].

(5) a. frumos-Ø băiat
beautiful boy

b. frumoas-e fete
beautiful girls

2. Romanian definite article has allomorphs for number and gender, however, I will focus on the masculine singular -(u)l.

3. Pre-nominal adjectives in Romanian are ambiguous between restrictive and non-restrictive readings while post-nominal adjectives have a restrictive reading (cf. Marchis & Alexiadou 2009, Petroj 2019, Petroj (to appear)).

(6) a. băiat frumos-Ø
boy beautiful
b. fete frumoas-e
girls beautiful

Next, Romanian plain adjectives are allowed post-nominally as adjectival predicates or predicates of relative clauses. In contrast, Romanian adjectives hosting D are not allowed in these positions, as illustrated in (7):

(7) Seba îi frumos/*frumos-ul (Rom)
Seba is handsome/handsome.the

3. Serbian NP

Serbian adjectives also agree in gender and number with the noun they modify, as in (8):

(8) a. *lep-Ø* *dečak*
beautiful.M.SG boy. M.SG
b. *lep-e* *devojke* (Srb)
beautiful.F.PL girls.F.P

Serbian does not have articles; however, it does have a way to express definiteness, namely, through different forms of adjectives: short (SF) and long (LF). These two forms are considered by some authors (Aljović 2002, Despić 2011, Talić 2013) to correspond to definite/specific and non-definite/non-specific interpretations respectively.

Table 1. Serbian adjectives – Lexical forms

	Serbian adjectives	
Masculine	nòv	nòv-i
Feminine	nóv-a	nòv-a:
	new.SF (non-definite/non-specific)	new.LF (definite/specific)

While Romanian adjectives can occur on either side of the noun, Serbian ones can only occur pre-nominally. There is, however, a restriction when it comes to their distribution: SF adjectives can be used both attributively and predicatively, whereas LF adjectives can generally be used only in the attributive position (cf. (9)

& (10))[4]. This will be important later in mixed structures, since Serbian adjectives carry this characteristic into CS, where the same restrictions apply, as illustrated by (11):

(9) zgodan/zgodni dečko (Srb)
handsome.SF/LF guy

(10) Seba je zgodan/*zgodni. (Srb)
Seba is handsome.SF/LF

(11) Seba **îi** *zgodan/*zgodni*. (CS)
Seba is handsome.SF/LF

4. Code-switched TNP

4.1 Distribution and issues

Before going into the analysis, Table 2 offers a summary and comparison of the elements (Romanian and Serbian) that participate in CS. The relevant constructions and interactions between the Romanian definite article -**(u)l**, and Serbian nouns and adjectives in definite contexts are highlighted in gray and repeated in (12)–(13).[5]

4. In some cases, a long-form adjective appears to occur predicatively, however, such cases have been convincingly analyzed in the literature as involving ellipsis, with the adjective modifying a null noun (see e.g. Bošković 2014, Talić 2017).

5. Although only definite contexts are considered in this paper, two important characteristics of Serbian adjectives appear in indefinite constructions in CS in (i) and (ii). First, (i) shows the short-form adjective can occur on either side of the noun. Secondly, although the long-form adjective is not allowed in either position in (ii) we will see later that long-form adjectives are not completely disallowed in CS, however, being specified for definiteness, they cannot occur in non-definite contexts.

(i) a. **un** *težak* *ispit*
a hard.SF exam

b. **un** *ispit* *težak*
a exam hard.SF

(ii) a. *__un__ *teški* *ispit*
a hard.LF exam

b. *__un__ *ispit* *teški*
a exam hard.LF

Table 2. Relevant characteristics of the Romanian and Serbian TNP elements

	Romanian	**Serbian**
Definite articles (D)	χ Agreement with N (gender & number)	χ
Nouns (N)	χ Phi-features: gender & number	χ Phi-features: gender & number
Adjectives (A)	χ 1 form Hosting definite articles: χ Occurring pre- and post-nominally Agreement with N (gender & number)	χ 2 forms: short (indefinite) & long (definite) Occurring pre-nominally Agreement with N (gender & number)

(12) a. *ispit*-**ul** *težak*
exam.the difficult.SF

b. **težak* *ispit*-**ul**
difficult.SF exam.the

c. **težak*-**ul** *ispit*
difficult.the exam

(13) a. *teški* *ispit*-**ul**
difficult.LF exam.the

b. **ispit*-**ul** *teški* (CS)
exam.the difficult.LF

There are several interesting points illustrated by the paradigm in (12)–(13). First, (12a) and (13a) show that Serbian nouns can act as hosts for definite articles, something that Serbian nouns do not do in Serbian (it lacking definite articles overall). This introduces important questions about the cliticization of the definite article in CS (but in Romanian, as well). Specifically, what is the mechanism that allows both Romanian and Serbian nouns to host definite articles? Additionally, can the mechanism proposed for Romanian account for CS constructions, as well? Moreover, what is particularly interesting here is that, in contrast to nouns, Serbian adjectives cannot host Romanian definite articles, although definite articles can be hosted by adjectives in Romanian. Additionally, with respect to their position, while Serbian adjectives can only occur pre-nominally in Serbian NPs, they can occur both pre- and post-nominally in CS, though their distribution is conditioned by the long/short form distinction. Examples in (12b) and (13a) show that LF adjectives are only allowed pre-nominally, and SF adjectives can only

occur post-nominally in definite contexts ((12a) vs. (12b)). However, in indefinite contexts, SF Serbian adjectives are allowed both pre- and post-nominally (see ft. 5). This has additional implications with respect to the analysis of Serbian adjectives in general, recognizing that there is a correlation between their predicative/attributive use in Serbian and their distribution with respect to their position and definiteness in CS examples. In other words, SF adjectives can be used both attributively and predicatively, whereas LF adjectives cannot be used predicatively at all (as an adjectival predicate in or as predicates of a relative clause) in Serbian, and this characteristic is transferred to CS, within the TNP environment. Finally, the 'double definiteness' that occurs in cases like (13a) is a deviation from both input grammars, given that neither Romanian nor Serbian expresses definiteness on more than one category within the same TNP (Romanian expressing it through the cliticization of the definite article on either the noun or the adjective, while Serbian expresses it through long-form adjectives). In the following sections, I will address the issues from above, and offer an account for the cliticization of definite articles in CS, which can also be applied to Romanian.

4.2 Description and analysis

There are three relevant elements that participate in the CS TNP: Romanian definite articles (D), Serbian nouns (N), and Serbian adjectives (A). The relevant paradigm from (12) and (13) is repeated below.

(12) a. *ispit***ul** *težak*
exam.the difficult.SF
b. **težak* *ispit*-**ul**
difficult.SF exam.the

(13) a. *teški* *ispit*-**ul**
difficult.LF exam.the
b. **ispit*-**ul** *teški* (CS)
exam.the difficult.LF

When it comes to D, here, I assume the following features for it: [val: +def], [uval: phi-features], [uval: nominal]. As a suffix, D requires a phonological host and this is the linearly closest element in the DP. However, recall that not all elements (e.g. adverbs) can act as hosts (cf. (5)). Importantly, while Romanian adjectives can host definite articles, Serbian adjectives cannot. To approach these issues, I will define the potential article host as in (1):

(1) *An element with a nominal feature that undergoes agreement with D.*

Nouns and adjectives both having nominal features, the inability of Serbian adjectives to host Romanian definite articles cannot be due to featural incompatibility,

but due to other factors that will be discussed below. Regarding Ns in CS, both Serbian and Romanian nouns are potential hosts for D, therefore, I assume the following matrix for nouns participating in CS: [uval: def], [val: phi-features], [val: +nominal]. There is, however, one curious feature of nouns in this variant of CS stemming from the NP/DP parameter setting. This is the unvalued definiteness feature that exists in Romanian, but should not be available to Serbian nouns in Serbian, since there is no D element in Serbian to value it. In order for the derivation to not crash, any noun in CS (whether Romanian or Serbian) has to undergo agreement with D. Here, I adopt Liceras et al. (2008) 'internal dominance' whereby, during CS, if a category differs feature-wise in the participating languages, the missing feature may be imported from the dominant language as a *grammaticized* feature. Liceras et al. (2008) discussed this for DP-internal CS in Spanish-English bilinguals; I assume something similar to occur here, where the unvalued [def] feature on the noun from Romanian is gramaticized and consequentially imported to Serbian nouns[6].

Moving on to adjectives, both Romanian and Serbian adjectives participate in CS. First, Romanian As have one lexical form and they can act as hosts for D. Second, Serbian ones have two lexical forms and both participate in CS, however, they cannot host D. Both forms of Serbian adjectives participating in CS introduces a new complexity; the restricted distribution of LF/SF Serbian adjectives that does not exist in Romanian. This additional definite/indefinite distinction, seemingly independent of the definite article (definite articles not existing in Serbian), needs to be accounted for in the mixed environment. Serbian adjectives undergo agreement with the noun, and I suggest that in CS, A may also undergo agreement with Romanian D, given that the occurrence of the two forms is conditioned by the presence or absence of the definite article, and by the position it occupies. Additionally, both Romanian and Serbian adjectives can occur pre- and post-nominally in CS, with certain restrictions: Romanian adjectives can only host definite articles if they occur pre-nominally; similarly, in definite contexts, Serbian adjectives can occur in their long (definite) form only pre-nominally, and in their short-form only post-nominally. Therefore, I will assume this feature matrix for adjectives participating in CS: [uval: def], [uval: phi-features], [uval: nominal]. A summary of the feature configuration for the three elements participating in CS discussed above is offered in Table 3:

6. Due to space limitations, the grammaticization of the [def] feature on Serbian nouns in CS will not be discussed here. For additional details, see Petroj (to appear).

Table 3. Feature configuration of elements participating in CS

D	A	N
[val: def]	[uval: def]	[uval: def]
[uval: phi-features]	[uval: phi-features]	[val: phi-features]
[uval: nominal]	[uval: nominal]	[val: nominal]

The status of adjectives in languages that have two positions available is very controversial (Ticio 2003; Cinque 2010, i.a.). Here, I am adopting the view from Cinque (2010), where pre-nominal adjectives are adverbial modifiers of the noun, and post-nominal adjectives are reduced relative clauses with a predicate adjective. This view applied here suggests that only SF adjectives can be found post-nominally, which is also the position for reduced relative clauses, and this is what CS examples show, as well. In contrast, Serbian LF adjectives cannot be used as reduced relative clauses, therefore, the only position where they may occur is pre-nominally. In other words, in principle, SF Serbian adjectives can occur wherever a plain Romanian adjective (i.e. an adjective not hosting D) occurs, and an LF Serbian adjective can only occur pre-nominally, which is the position where Romanian adjectives hosting D occur. I will also assume that modifying adjectives are NP-adjoined (see Petroj, to appear).

5. Towards an analysis

When the structures of two languages mix, elements from both languages have to adapt to a new environment, and elements with different properties find a way to interact with each other. The most striking consequence of this is illustrated in (14), a case of 'double definiteness' with an LF Serbian adjective and the cliticization of D to the noun. Recall that, while Serbian only expresses definiteness through LF adjectives, Romanian can express definiteness on nouns and adjectives by the cliticization of the definite article on either element (but not on both simultaneously).

(14) *teški* *ispit*-**ul**
difficult.the exam.the

To derive the CS paradigm from (12)–(13), I offer here an initial account that combines Multiple Agree (Hiraiwa 2001) and Affix Hopping (Chomsky 1957, also known as PF merger)[7]. While Affix Hopping will be used in merging the affix **-(u)l**

7. The inability of articles being hosted through a more traditional mechanism, such as N-to-D movement (Longobardi 1994, Ungureanu 2006, i.a.) in Romanian or CS structures is

with a host in PF, Agree will be used to determine what kind of element can be considered a host for -**(u)l**. Specifically, I assume that a host is nominal element that undergoes agree with D. In CS, I propose that -**(u)l** is an affix in D, valued for definiteness, and that it hops on a linearly closest potential host in PF. Since adjacency matters for Affix Hopping, it is important to note that some elements are skipped in CS, such as adverbs in Romanian (4) and Serbian adjectives as in (14). If a potential host is determined by Agree, the adverb is immediately excluded since it does not undergo Agree with D. As far as the inability of Serbian adjectives to host D, I argue that, since adjectives in Serbian have a lexically definite form (LF), it cannot and need not express definiteness twice. Regarding why Affix Hopping can skip adverbs and adjectives, following Bobaljik (1994) and Ochi (1999), I assume that adjuncts do not interfere with the adjacency requirement for Affix Hopping, so the reason why -**(u)l** can skip elements in (4) and (14) is because adverbs and adjectives are adjuncts in CS.[8]

Regarding the interaction of D, N, and A in CS, I assume that both Romanian and Serbian As have unvalued definiteness, phi-, and nominal features. They agree with N (which values their phi- and nominal features) and with D (which values their [def] feature). The valuation of the [def] feature on the Serbian A has a morphological reflex; the morphological realization of this valuation results in the long-form of the adjective. D is then involved in two processes:

1. *D must undergo (multiple) Agree*;
2. *D must be hosted by an element it Agrees with*.

The derivation for (14) is given below:

Step 1 – **Agree**: A and N undergo agreement, N valuing the unvalued nominal and phi-features on the pre-nominal Serbian adjective (it is possible that A and N undergo feature sharing for [uval: def]).

težak	*ispit*	➡	*težak*	
difficult	exam		difficult	
[uval: def]	[uval:def]		[uval: def]	
[uval: phi features]	⟵		[val: phi-features]	**[val: SG, M]**
[uval: nominal]	⟵		[val: +nominal]	**[uval: nom]**

shown through independent evidence from Left Branch Extraction (LBE) cases discussed in Petroj (2019) and Petroj (to appear).

8. Romanian adverbs are just like adverbs in English. Namely, in English, negation 'not' blocks affix hopping of T elements on the verb, as in (i), but adverbs do not, as in (ii):

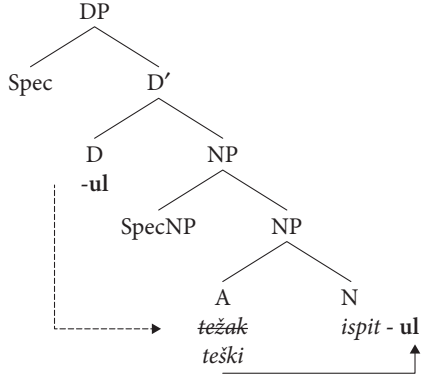
(i) *John NOT walk-ed to school.
(ii) John always walk-ed to school.

Step 2 – Agree: D undergoes multiple agree with A and N. This way it has its unvalued nominal and phi-features valued. In return, D values the unvalued [def] feature on the A as definite, the long-form being the morphological realization of an adjective that is valued as definite (recall that adjectives valued for definiteness can only occur pre-nominally in CS). It also values the unvalued [def] feature of the noun. The results of these valuations are given below:

D:	**-(u)l**	➡	**-(u)l**
	the		the
	[val: +def]		[val: +def]
	[uval: phi features]		**[val: SG, M]**
	[uval: nominal]		**[val: nominal]**
A:	*težak*	➡	*teški*
	difficult.SF		difficult.LF
	[uval: def]		**[val: +def]**
	[uval: SG.M]		[val: SG, M]
	[uval: nominal]		[val: nominal]
N:	*ispit*	➡	*ispit*
	exam		exam
	[uval: def]		**[val: def]**
	[val: SG, M]		[val: SG, M]
	[val: +nominal]		[val: nominal]

Step 3 – Affix Hopping: In PF, **-(u)l** needs a host to hop on. I propose that Serbian adjectives are not adequate hosts for D since Serbian adjectives already have inflection through which definiteness is expressed. This is the long form, which is the morphological realization of the [def] feature valuation between D and the adjective. D then still needs a host; D then hops onto the noun and thus becomes realized on it.

(15)



The 'double definiteness' occurs in CS due to the failure of D to fulfill its Affix Hopping requirement by the first element that it undergoes Agree with. Notice that the

reason why (13) is ungrammatical is because D could not have agreed with the noun, valuing its [def] feature (this would make it a proper host for - **(u)l**, without also agreeing for [def] with the adjective, valuing it as [def], which then requires the long form.

(13) **težak* *ispit*-**ul**
heavy.SF computer.the

I argue that in (12), the adjective is actually part of a reduced relative clause (Cinque 2010), hence too deeply embedded to agree with D.

(12) *ispit*-**ul** *težak*
exam.the difficult.SF

6. Constraints on CS: A brief discussion[9]

There have been several CS constraints that ban word-internal CS. These include (1) the Free Morpheme Constraint (Poplack 1980), which claims that only free morphemes can be CS; (2) The Functional Head Constraint (Belazi, Rubin, & Toribio 1994), which bans CS of functional heads; and (3) the PF Disjunction Theorem (MacSwan 1999), which claims that CS below an X^0 is not possible. Given that Romanian D is a bound morpheme which can be hosted by Serbian nouns, I argue that these constraints do not hold crosslinguistically in CS. In fact, the data presented here supports Rhao and den Dikken's (2003) reformulated PF Disjunction Theorem. They claim that word-internal CS is allowed, so long as the two elements from different languages form a phonological, not a syntactic head. As seen in the previous section, Romanian D and the Serbian noun form a phonological, not a syntactic head, since Affix Hopping is a mechanism that applies in PF.[10]

9. One of the most well-known models regarding the analysis of CS utterances is Myers-Scotton's (1993) Matrix Language Frame Language Frame (MLF) model which will not be discussed here. Specifically, following MacSwan (2005), when approaching Romanian-Serbian CS from the perspective of generative syntax, the model shows many inconsistencies with respect to its crosslinguistic application to various language pairs. In addition, Petroj (2019, to appear) also shows that CS is contextually conditioned relevant to clauses and phasal domains, and, as such, assuming a dominant and an embedded language in the style of Myers-Scotton (1993) would prove futile and inaccurate.

10. Due to space limitations, I am unable to further discuss these constraints here. For a more comprehensive discussion, data, and analysis, I refer the reader to Petroj (to appear).

7. Conclusion

I investigated the structure of the TNP in an unusual environment that has previously not been explored: in Romanian-Serbian CS. Apart from exploring the mechanisms that a mixed language system makes use of in combining two clashing parameters, the nature of code-switching allows for the investigation of relevant elements belonging to two different languages detached from their input grammar. By looking at these elements in unknown environments, we can discover properties that were previously impossible to approach. By investigating the distribution of the definite article on elements belonging to a language without definite articles, I demonstrated that through CS research, new perspective may arise towards analyzing controversial phenomenon in languages in isolation. Finally, I showed that, contrary to claims by Poplack (1980), Belazi et al. (1994), and MacSwan (1999), and following Bandi-Rao and den Dikken's (2014), TNP-internal CS is allowed in this language pair.

Acknowledgments

I would like to thank Dr. Željko Bošković, Dr. Diane Lillo-Martin, and Dr. Ian Roberts for their valuable discussions and insights. I am also grateful to the audiences at the CUNY Syntax Supper seminar and at LSRL 47 where this work was presented. Additionally, I appreciate all the constructive criticism and suggestions made by the anonymous reviewer for this article. Finally, a huge thanks to my informants for their patience and availability. All errors are mine.

References

Abney, Steven P. 1987. "The English Noun Phrase in its Sentential Aspects." Doctoral Dissertation, MIT.

Aljović, Nadira. 2002. "Long Adjectival Inflection and Specificity in Serbo-Croatian." *Recherches Linguistiques de Vincennes* 31: 27–42. https://doi.org/10.4000/rlv.351

Baker, Mark C. 2003. *Lexical Categories: Verbs, Nouns, and Adjectives* (Studies in Linguistics 102). Cambridge University Press. Cambridge. https://doi.org/10.1017/CBO9780511615047

Bandi-Rao, Shoba, and Marcel den Dikken. 2014. "Light Switches. On v as a Pivot in Codeswitching, and the Nature of the Ban on Word-Internal Switches." In *Grammatical Theory and Bilingual Codeswitching*, ed. by J. MacSwan, 161–184. Cambridge, MA: MIT Press.

Belazi, Hedi M., Edward J. Rubin, and Almeida J. Toribio. 1994. "Code switching and X-bar theory: The functional head constraint." *Linguistic Inquiry* 25: 221–237.

Bobaljik, Jonathan. 1994. "What does adjacency do? *MIT Working Papers in Linguistics* 22: 1–32.

Bošković, Željko. 2008. "What Will You Have, DP or NP?." *Proceedings from NELS 37: North East Linguistic Society*. 101–114.

Bošković, Željko. 2012. "On NPs and Clauses." In *Discourse and Grammar: From Sentence Types to Lexical Categories*, ed. by G. Grewendorf, and T. E. Zimmermann, 179–242.
Bošković, Željko. 2013. "Adjectival Escapades." *Proceedings from FASL 21: Formal Approaches to Slavic Linguistics*. 1–25.
Bošković, Željko. 2013. *Getting really edgy: On the edge of the edge*. Ms. University of Connecicut.
Bošković, Željko. 2014. "Now I'm a Phase, Now I'm Not a Phase: On the Variability of Phases with Extraction and Ellipsis." *Linguistic Inquiry* 45: 27–89. https://doi.org/10.1162/LING_a_00148
Bowers, John. 1991. "The Syntax and Semantics of Nominals." In *Cornell University Working Papers in Linguistics* 10: 1–30.
Chierchia, Giuseppe. 1998. "Reference to Kinds across Languages." *Natural Language Semantics* 6: 339–405. https://doi.org/10.1023/A:1008324218506
Chitoran, Ioana. 2000. *The phonology of Romanian: A constraint-based approach*. New York: Mouton de Gruyter.
Chitoran, I. 2002. In *The Phonology of Romanian: A Constraint-Based Approach*, ed. by Van der Hulst, H. van Riemsdijk, and J. Koster. NY: Mouton de Gruyter. https://doi.org/10.1515/9783110889185
Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton de Gruyter.
Cinque, Gugliermo. 2010. *The Syntax of Adjectives: A Comparative Study*. Linguistic Inquiry Monographs. MIT Press. https://doi.org/10.7551/mitpress/9780262014168.001.0001
Corver, Norbert. 1990. "The Syntax of Left Branch Extractions". Doctoral Dissertation, Tilburg University.
Despić, Miloje. 2011. "Syntax in the Absence of Determiner Phrase". Doctoral dissertation, University of Connecticut.
Dobrovie-Sorin, Carmen. 1994. *The Syntax of Romanian. Comparative Studies in Romance*, ed. by J. Koster, and H. van Riemsdijk. New York: Mouton de Gruyter. https://doi.org/10.1515/9783110886597
Dobrovie-Sorin, Carmen, and Ion Giurgea (Eds.). 2013. *A Reference Grammar to of Romanian*. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/la.207
Fukui, Naoki. 1988. "Deriving the Differences Between English and Japanese." *English Linguistics* 5: 249–270. https://doi.org/10.9793/elsj1984.5.249
Halpern, Aaron. 1992. *On the Placement and Morphology of Clitics*. Center for the Study of Language and Information: Dissertations in Linguistics
Hiraiwa, Ken. 2001. "Multiple Agree and the Defective Intervention Constraint in Japanese." In *The Proceedings of the MIT-Harvard joint conference (HUMIT 2000), MITWPL* 40, ed. by O. Matushansky, 67–80. Cambridge, MA: MITWPL.
Liceras, Juana. M., Raquel Fernández Fuertes, Susana Perales, Rocio Pérez-Tattam, and Todd, K. Spradlin 2008. "Gender and Gender Agreement in Bilingual Native and Non-native Grammars: A View from Child and Adult Functional-Lexical Mixings." *Lingua* 118: 827–851. https://doi.org/10.1016/j.lingua.2007.05.006
Longobardi, Guiseppe. 1994. "Reference and Proper Names." *Linguistic Inquiry* 25: 609–665.
MacSwan, Jeff. 1999. "A Minimalist Approach to Intrasentential Code Switching: Spanish–Nahuatl Bilingualism in Central Mexico". Doctoral dissertation, UCLA.
MacSwan, Jeff. 2005. "Codeswitching and Generative Grammar: A Critique of the MLFM Model and Some Remarks on Modified Minimalism." *Bilingualism: Language and Congition* 8 (1): 1–22. https://doi.org/10.1017/S1366728904002068

Marchis, Michaela, and Artemis Alexiadou. 2009. "On the distribution of adjectives in Romanian." In *Romance Languages and Linguistic Theory: Selected papers from 'Going Romance' Amsterdam 2007*. 161–178. https://doi.org/10.1075/rllt.1.09mar

Myers-Scotton, Carol. 1993. *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon Press

Ochi, Masao. 1999. "Multiple Speel-Out and PF-Adjacency." *Proceedings from NELS 29: North East Linguistic Society*, 293–306.

Pereltsvaig, Asya. 2007. "On the Universality of DP: A View from Russian." *Studia Linguistica* 61: 59–94. https://doi.org/10.1111/j.1467-9582.2007.00129.x

Petroj, Vanessa. 2019. "Extract to Unravel: Left Branch Extraction (LBE) in Romanian-Serbian Code-Switching." *Proceedings from FDSL 12: Formal Approaches to Slavic Linguistics*.

Petroj, Vanessa. to appear. "Ignoring Language Barriers: Romanian-Serbian Code-Switching." Doctoral dissertation, University of Connecticut.

Poplack, Shana. 1980. "Sometimes I'll Start a Sentence in Spanish y Termino Espanol: Toward a Typology of Code-Switching." *Linguistics* 18: 581–618. https://doi.org/10.1515/ling.1980.18.7-8.581

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.

Talić, Aida. 2013. "Extraordinary Complement Extraction (Handout)." Presented at LSALAA 2013. Paris, France.

Talić, Aida. 2017. "From A to N and Back: Functional and Bare Projections in the Domain of N and A." Doctoral Dissertation.

Ticio, Emma M. 2003. "On the structure of DPs." Doctoral dissertation, University of Connecticut.

Tomić Olga M. (Ed.). 2004. *Balkan Syntax and Semantics*. Amsterdam: John Benjamins.

Tomić, Olga M. 2006. *Balkan Sprachbund Morpho-Syntactic Features*. Dordrecht: Springer. https://doi.org/10.1007/1-4020-4488-7

Ungureanu, Mona-Luiza. 2006. "DP-internal structure and movement in Romanian." Doctoral dissertation, McGill University.

Split auxiliary selection with affected subjects in Old Majorcan Catalan

Sebastià Salvà i Puig
Centre de Lingüística Teòrica, Universitat Autònoma de Barcelona

We aim to shed light on the split auxiliary selection found in Old Majorcan Catalan, in constructions where, instead of *haver* ('have'), *esser* ('be') is used as the auxiliary verb for compound tenses, although just with persons 1 & 2; not only with unaccusative verbs, but also with transitive and unergative verbs. We claim that, in fact, these constructions have affected or undergoer subjects, derived from a lower position: at least, [Spec, ProcessP]. Hence, this system for auxiliary selection is mixed: both person-driven and event-driven. The Majorcan data provide evidence in favour of Ramchand's (2008; 2018) decomposition of the *v*P, and for a nanosyntactic lexicalization of the auxiliaries: *haver* lexicalizes a chunk such as [Aux, *i**], while *esser* lexicalizes only [Aux], in case a Person$_{[+participant]}$ head intervenes.

Keywords: split auxiliary selection, person-driven auxiliary selection, affected subject, undergoer, unaccusativity, possessive/interest relationship, subject of result construction, nanosyntactic lexicalization, Majorcan Catalan, Northern Catalan

1. Introduction

In Catalan, split auxiliary selection – that is to say, the use of the auxiliary verb *esser/ésser/ser* ('be') with persons 1 & 2 of compound tenses – has been attested in two different ways:

i. In a widespread or generalised manner (with all kind of verbs):
   a. in current Northern Catalan or *rossellonès* (spoken in Rosselló, Vallespir, Conflent and Capcir) (Alcover 1903; Veny & Massanell 2015; Massanell 2017);
   b. in some North-Eastern varieties of Central Catalan (in the south of Alt Empordà, Garrotxa and Ripollès; and in Baix Empordà, Pla de l'Estany, Gironès, Selva, Plana de Vic and Lluçanès) (Puig i Sais 1906; Badia 1981; Rigau 1998).

https://doi.org/10.1075/rllt.16.14sal

ii. Sporadically (optionally, not always, and just with some verbs):
   a. in Transitional Northern Catalan (in the north of Alt Empordà, Garrotxa and Ripollès; and in Alta and Baixa Cerdanya) (Veny & Massanell 2015);
   b. in Majorcan Catalan.

We are going to focus on Majorcan Catalan: particularly, on the variety spoken by elder people from some villages (Llucmajor, Montuïri, Santanyí, Felanitx…).

In current or modern Majorcan Catalan – in the most widespread variety – the auxiliary verb is always *haver* ('have'), for all kinds of verbs and constructions, for all compound tenses (present perfect, past perfect…), for all moods (indicative or optative) and for all grammatical persons. So current Majorcan Catalan behaves like standard Catalan, with one peculiarity: in several cases, *past participle agreement* (PPA) with the internal argument is possible (Rosselló 2002; Salvà 2015, 2017, 2018).[1]

However, the auxiliary verb *esser* ('be') can be used for compound tenses in the archaic variety spoken by elder people from some villages (Llucmajor, Montuïri, Santanyí, Felanitx...) in unaccusative, passive and reflexive constructions (1b–d), with some peculiarities:

i. *Esser* is more frequent with persons 1 & 2 (especially of the singular), and mainly in the present perfect tense of indicative.
ii. By contrast, person 3 (and also 1 & 2 in other tenses/moods) tends to select *haver* with non-reflexive unaccusative verbs; but *esser* is preferred with reflexive and passive constructions – probably, according to Ramos (2005), because they possess an explicit morphological mark for the absence of external argument: either the reflexive clitic *se*, or the passive auxiliary verb *esser* + a passive past participle.

---

1. Majorcan Catalan is one of the Romance varieties most prone to PPA in compound tenses. According to Rosselló (2002: 1932–1933), PPA is possible especially with (feminine) 3rd person accusative clitics (mandatorily with *la*: CL.3FEM.SG), partitive *ne/en/n*' included, but also with 1st and 2nd person accusative clitics, and in cases of *wh*-movement – here, we can add contrastive focus too – and when the internal argument is promoted to grammatical subject – passive, reflexive and unaccusative constructions (especially with motion verbs of inherent direction). Moreover, in some cases, PPA is even possible when the direct object remains *in situ*, in its canonical post-verbal position – *e. g.*, (i) – (Salvà 2015, 2017, 2018), although this property is getting lost among the youngest speakers.

(i) *S'* *ha* *{rentades/rentat}* *ses* *mans.*
CL.REFL has washed.{FEM.PL/MAS.SG} the.FEM.PL hands.FEM.PL
'{He/she} has washed {his/her} hands'

(1) a. Unergative/transitive verbs (*e. g., empènyer* 'push'):
he empès 'I have pushed'
has empès 'you have pushed'
ha empès '{he/she} has pushed'
hem empès 'we have pushed'
heu empès 'you have pushed'
han empès 'they have pushed'

b. Unaccusative verbs (*e. g., arribar* 'arrive'):
***som** arribat*; ***som** {arribada/*arribat}* 'I have arrived'
***ets** arribat*; ***ets** {arribada/*arribat}* 'you have arrived'
ha arribat; *ha {arribat/arribada}* 'he/she has arrived'
hem {arribat/arribats}; *hem {arribat/arribades}*
?***som** {arribats/*arribat}*; ?***som** {arribades/*arribat}*
heu {arribat/arribats}; *heu {arribat/arribades}*
?***sou** {arribats/*arribat}*; ?***sou** {arribades/*arribat}*
han {arribat/arribats}; *han {arribat/arribades}*

c. Reflexive constructions (*e. g., pentinar-se* 'comb one's hair'):
*me **som** pentinat*; *me **som** {pentinada/*pentinat}*
*t'**ets** pentinat*; *t'**ets** {pentinada/*pentinat}*
*s'**és** pentinat*; *s'**és** {pentinada/*pentinat}*
*mos **som** {pentinats/pentinades/*pentinat}*
*vos **sou** {pentinats/pentinades/*pentinat}*
*se **són** {pentinats/pentinades/*pentinat}*

d. Passive constructions (*e. g., esser segrestat* 'be kidnapped'):
***som** estat segrestat*; ***som** {estada segrestada / *estat segrestat}*
***ets** estat segrestat*; ***ets** {estada segrestada / *estat segrestat}*
***és** estat segrestat*; ***és** {estada segrestada *estat segrestat}*
***som** estats segrestats*; ***som** estades segrestades*
***sou** estats segrestats*; ***sou** estades segrestades*
***són** estats segrestats*; ***són** estades segrestades*

Notice that, in (1b–d), with the auxiliary verb *esser*, PPA with the internal argument (here the grammatical subject) is always mandatory.

Furthermore, a fact that has been understudied until now is that the old Majorcan speakers with the archaic system in (1) – and even younger people, in some stereotyped cases, like the ones in (2) – *sporadically* use (optionally alternating with *haver*) the auxiliary verb *esser* in some transitive sentences (with an explicit direct object) – (2) and (3a–e) – and in some unergative sentences (3f):

(2) a. *La* ***som*** *vista.*
CL.ACC.3FEM.SG am seen.FEM.SG
'I've seen her'

b. ***Som*** *agafada una nirviada.*
am taken.FEM.SG a nervous-ness.AUGM.FEM.SG
'I've got so nervous'

c. *Bona la* ***som*** *feta!*
good.FEM.SG CL.ACC.FEM.SG am made.FEM.SG
'I've made a {mess / terrible mistake}!'
[oral examples from Llucmajor (Mallorca)]

(3) a. *Tomeu, que no m'* ***ets*** *sentit?*
Tomeu EVID.that not CL.1SG are.2SG heard
'Tomeu, haven't you heard me?'

b. *Des que* ***som*** *comprat un ase…*
since that am bought a donkey
'Since I've bought a donkey...' (or 'Since I possess a donkey that I bought…')

c. *Li* ***som*** *fet mal de cap, a sa padrina.*
CL.DAT.3SG am done head-ache MARK.DAT the grandmother
'I've given her a headache (to my grandmother)'

d. *Amor, com més pena m'* ***ets*** *causada,*
darling as more pain.FEM.SG CL.1SG are.2SG caused.FEM.SG
més voluntat t' he tenguda.
more will.FEM.SG CL.2SG have.I had.FEM.SG
'Darling, the more pain you've caused to me, the more I've wished you'

e. *Sa soca li* ***ets*** *danyada, a*
the trunk.FEM.SG CL.DAT.3SG are.2SG damaged.FEM.SG MARK.DAT
s' arbre.
the tree
'You've damaged its trunk (the tree's trunk)'

f. *Davall aquella figuera hi* ***som*** *festejat molts pics.*
under that fig-tree CL.LOC am courted many times
'Under that fig-tree, I've courted many times'
[examples from *Cançoner popular de Mallorca*, by Rafel Ginard]

Again, *esser* is used especially with 1SG & 2SG of indicative present perfect – although not exclusively, as shown by (4), in optative past perfect:[2]

2. The optative past perfect in (4) has a *reproachative* or *counterfactual* use, similar to the so-called *retrospective imperative* (*e.g.: haver moguts es dits* 'you should have moved your fingers'), in order to express the reproof for not having carried out an action that should have been accomplished.

(4) *Com es vent era a sa flauta,* ***fosses*** *remenats*
when the wind was in the flute be.PAST.OPT.2G stirred.MAS.PL
es dits.
the fingers.MAS.PL
Lit.: 'When the wind was through the flute, you should have stirred your fingers' (= 'It is when I was {there for you / in love with you} that you should have given me a signal, not now'
[example from the entry *esser* in *Diccionari català-valencià-balear*, by Alcover & Moll, *apud* Massanell (2017)]

As shown by (i) in fn. 1, (2), (3d–e) and (4), in Majorcan Catalan, PPA within transitive constructions is established with the direct object (not with the subject), even with the auxiliary verb *esser*.[3] This supports the idea that PPA and auxiliary selection are different phenomena, although related topics (Loporcaro 1998).

2. Previous accounts

Prescriptively, this use of *esser* with non-unaccusative verbs has been considered "very curious" and "weird", and hence, "abusive", "incorrect", "anti-grammatical" and "absolutely inadmissible" by traditional grammars (Alcover 1903; Obrador 1906a, 1906b; Fabra 1922; Moll 1952: § 481, 1975: § 159).

Descriptively (and focusing on Northern and North-Eastern varieties of Catalan), only Badia (1981: § 176; 1994: 561) notices that this use is limited to a reduced number of verbs (*e.g., veure* 'see'). By contrast, other authors point out

3. By contrast, in the other Catalan dialects pointed out at the beginning of § 1, the past participle remains invariable (with no agreement with the direct object) in similar cases (ii.a–c), except with an ACC.FEM clitic (ii.d):

(ii) a. *No* ***sou*** *vist la processó.*
not are seen the procession.FEM.SG
'You've not seen the (religious) procession'

b. ***So*** *posat les claus damunt la taula.*
am put the keys.FEM.PL on *the table*
'I've put the keys on the table'

c. ***Soc*** *sofert molts desenganys.*
am suffered many disappointments.MAS.PL
'I've been disappointed many times'

d. *Jo ja la* ***som*** *digerida.*
am already CL.ACC.FEM.SG am digested
'I've already digested it'

that these dialects display *person-driven auxiliary selection* «with all kind of verbs», so independently of the event/argument structure (Alcover 1903; Puig i Sais 1906; Rigau 1998). Today, this is maybe the case for strictly Northern Catalan or *rossellonès*, which has generalised or extended the use of *ésser* with persons 1 & 2 to all kinds of verbs or constructions (Veny & Massanell 2015; Massanell 2017); and the same stands for some North-Eastern varieties of Central Catalan (Rigau 1998 & p. c.). Nevertheless, this conclusion does not seem to be accurate for Old Majorcan Catalan (nor for Transitional Northern Catalan), which display(s) sporadic and optional split auxiliary selection, often alternating with *haver.*

3. Our provisional analysis: The initiator-undergoer chain

If we pay attention to the sentences in (2) and (3), an *unintentional* or *non-volitional* reading is often available. If this is so, these Majorcan Catalan varieties would be placed quite low in Sorace's (2000, 2004) *auxiliary selection hierarchy* (ASH):[4]

(5)	TELIC CHANGE OF LOCATION >	***core unaccusatives:*** **select 'BE' (least variation)**
	TELIC CHANGE OF STATE >	cut-off point (French)
	ATELIC CHANGE OS STATE >	cut-off point (German/ Dutch/Danish)
	CONTINUATION OF PRE-EXISTING STATE >	
	EXISTENCE OF STATE >	cut-off point (Italian)
	UNCONTROLLED PROCESS >	cut-off point (Old Majorcan Catalan)???
	MOTIONAL CONTROLLED PROCESS >	
	NON-MOTIONAL CONTROLLED PROCESS	***core unergatives:*** **select 'HAVE' (least variation)**

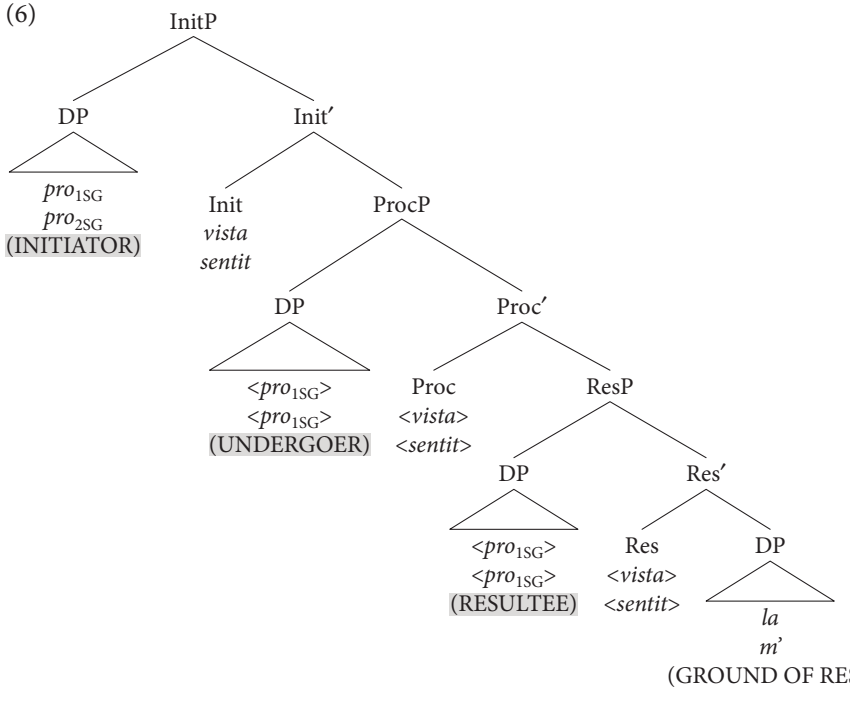
4. "Verbs at the BE end of the ASH are core unaccusatives and denote telic change; verbs at the HAVE end are core unergatives and denote agentive activity in which the subject is unaffected. Intermediate verbs between the two extremes incorporate telicity and agentivity to lesser degrees, and tend to have a less specified (basically stative) event structure" (Keller & Sorace 2003: 60) (underlining added).

Mateu (2002, 2009) translates Sorace's lexical semantic classes into 5 formal combinations: [[+T] [+r]] (telic change of location/state), [[+T][−r]] (atelic change of location/state), [[−T][−r]] (continuation of a pre-existing state / existence of state), [−R] (non-volitional internal cause) and [+R] (volitional internal cause).

If we look more carefully at the Old Majorcan sentences in (2)–(4), a common pattern can be established: their grammatical subject is *affected* (it is an *experiencer*, a *possessor* or an *interested* argument). A technical way of analysing this is appealing to Ramchand's (2008) event structure, with a head Process (between Initiation and Result) whose specifier is interpreted as an *undergoer* (as it undergoes or suffers the process of the event).

Take, for instance, (2a) and (3a). *Veure* 'see' and *sentir* 'hear' are stative verbs (Kiparsky 1998; Salvà 2017: § 4) and, according to Ramchand (2008), they have the simple configuration of an InitiationP (or StateP), and that is all. The specifier is interpreted as the holder of that state and the complement is interpreted as the rheme. However, as claimed by Jaque (2014) and Salvà (2017), some stative verbs – the ones that Jaque calls "high" (or "level-1") pure (or Kimian) stative verbs –, in some contexts (as present perfect), can behave as achievements and, therefore, unfold a whole event structure (that is: also ProcessP and ResultP). In this case, the subject argument is, at the same time, an initiator, an undergoer and a resultee:

(6)

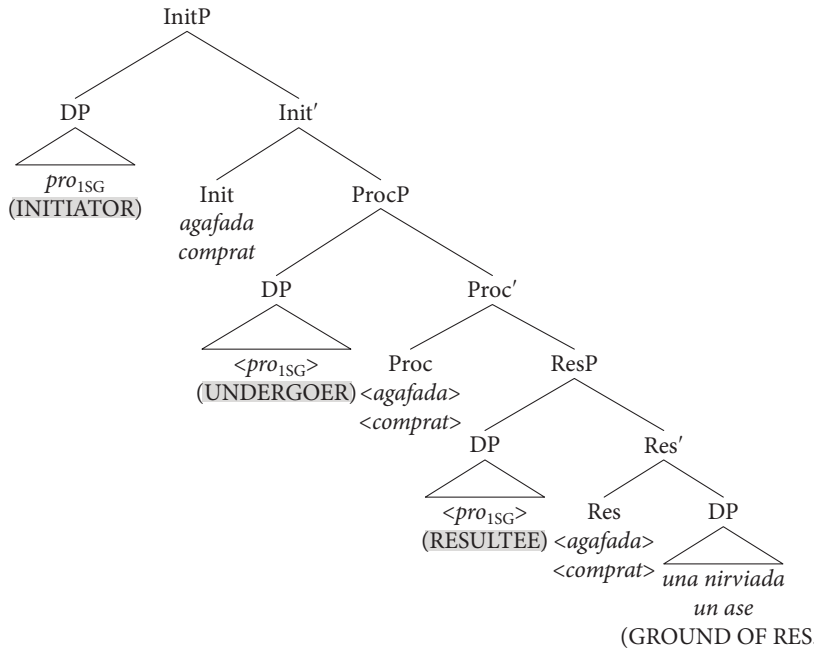


A similar analysis (7) can be proposed for the sentences (2b) and (3b), which express also achievements.

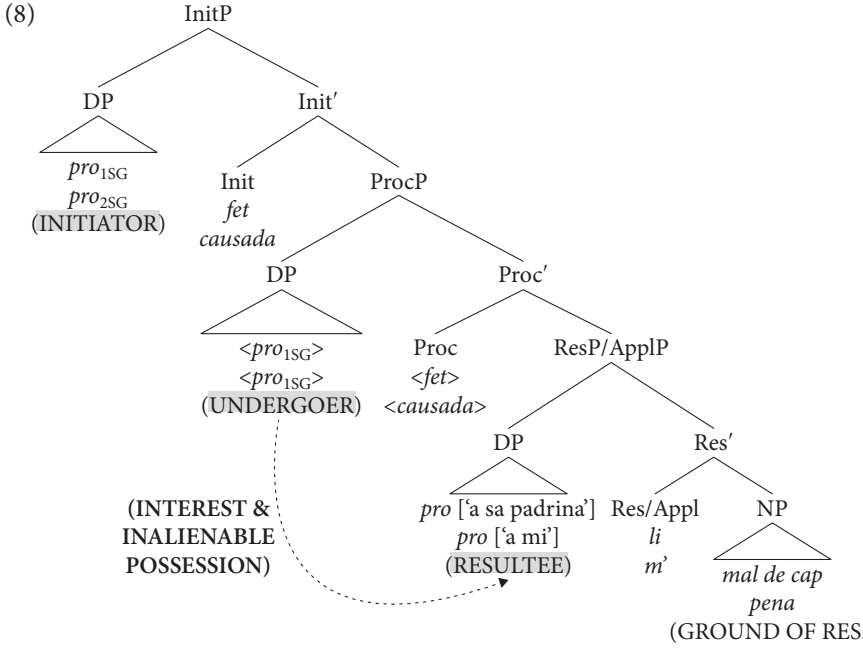
As for (3c–d), it is not so obvious that the subject be affected (an undergoer), but indeed it is, as it maintains an *inalienable possessive relationship* with the

resultee argument ('my grandmother' or 'me, your lover') – see the analysis in (8). A clearer case would be the one in (4), where the subject is the possessor of a part of his own body (*es dits* 'your fingers') – see the analysis in (9).

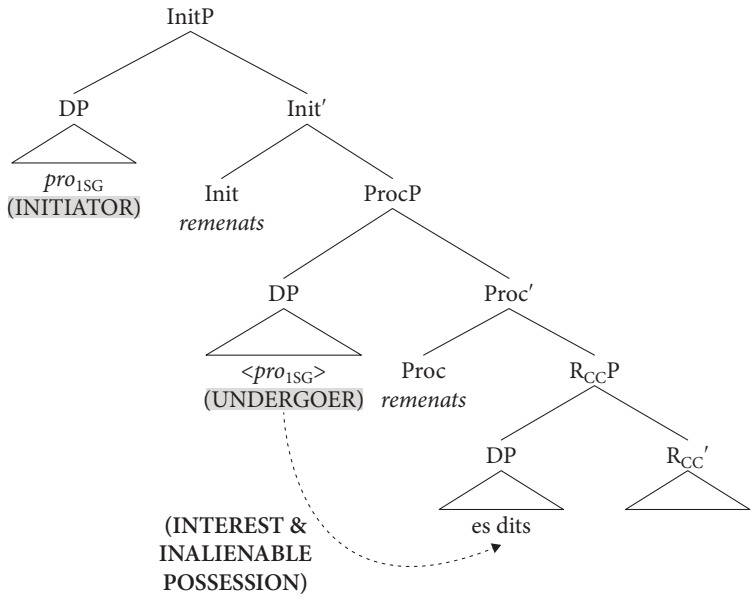
(7)



(8)



(9)

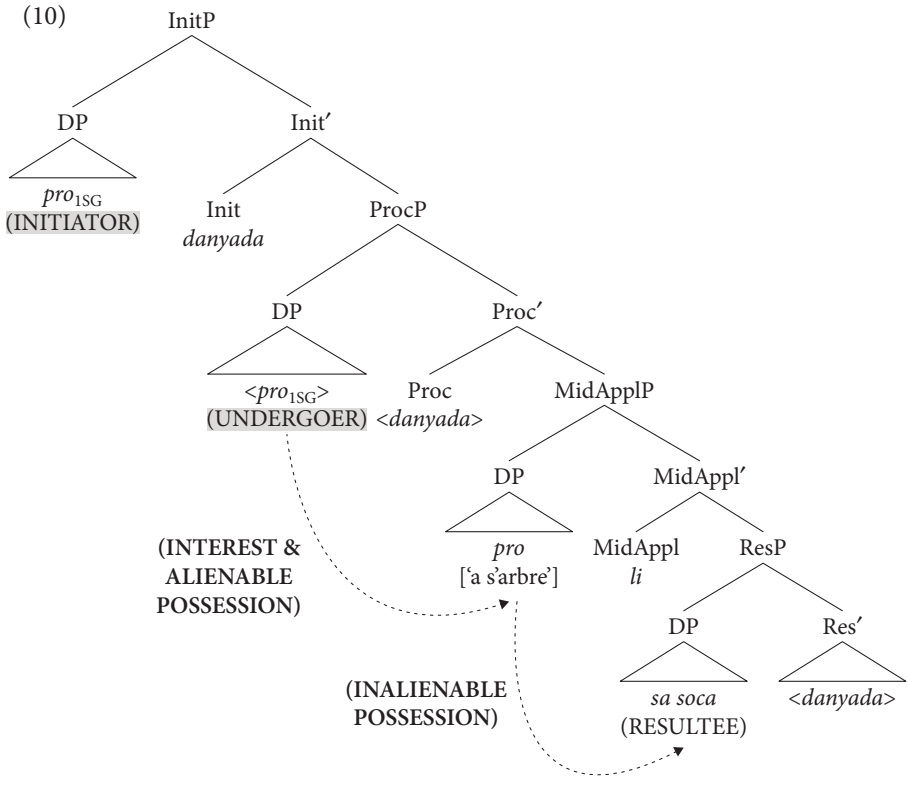


The most complex case would be (3e), which displays an extra argument. Following Cuervo (2008), here we have an *affected* or *middle Appl* over ResP, and the specifier of this MidApplP (*s'arbre* 'the tree') already maintains an inalienable possessive relashionship with the resultee ([Spec, ResP]: *sa soca* 'the trunk'). However, this is not an obstacle for the subject (in [Spec, InitP] and in [Spec, ProcP]) be also considered affected (an undergoer), if it can be conceived as being *interested* in the event – for instance, if he is the owner of the tree (in this case, there would be an *alienable* possessive relationship between [Spec, ProcP] and [Spec, MidApplP]) – see the analysis in (10).

Therefore, instead of using *high (ethic dative) applicatives* (between ProcP and InitP), Old Majorcan Catalan allows to Merge any *affected* or *interested* argument *directly in [Spec, ProcP]*, especially if it maintains a *possessive relationship* with another argument.

To sum up, we are dealing with *"unaccusative" constructions in a broad sense*,[5] with *affected subjects* derived from a lower position. In the Old Majorcan system of auxiliary selection, in order for *esser* ('be') to appear, [Spec, InitP] – if present – must form a *chain* with (at least) [Spec, ProcP] (where the argument is interpreted as an *undergoer*).

5. *Stricto sensu*, Ramchand (2008) considers as (proper) unaccusative structures those ones without an Init head (*e. g., The ice melt*), which correspond to Hale & Keyser's (2002) *core* unaccusatives.



As stated by Bentley (2016), in the transition from late Latin to early Romance, the establishment of the HABERE/ESSE alternation in the perfect marks a split (semantically motivated, by an *active/inactive* distinction) between two kinds of *intransitive* constructions: one with an S_A ("actor subject") and one with an S_U ("undergoer subject"). The former construction aligns with transitive constructions (with an A subject). Since A/S_A is the default choice for subject in terms of nominative/accusative alignment, *the perfect auxiliary* ESSE *indicates the marked choice for subject in Romance.* Starting from this, we claim that, in Old Majorcan Catalan, a new split emerges within *transitive* constructions, whose subject can be either a simple *initiator* or an *initiator-undergoer.*

4. Fulfilled predictions

Actually, any of the verbs that Ramchand (2008) considers to have an undergoer co-referent with the initiator (so, an argument with these two event roles at the same time) could be used with the auxiliary verb *esser*.

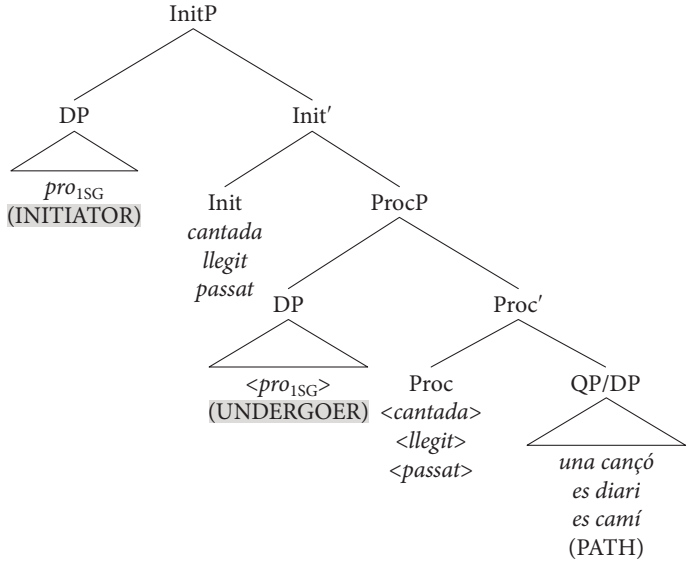
This is what happens with creation/consumption verbs, and with any other verb with a *path* (or *incremental theme*) direct object:

(11) a. *Som cantada una cançó.*
am sung.FEM.SG a song
'I've sung a song'

b. *Som llegit es diari.*
am read the newspaper
'I've read the newspaper

c. *Som passat es camí.*
am gone.through the path
'I've gone through the path'

(12)



And the same would stand with verbs like *trobar* 'find' and *perdre* 'lose' (13), whose initiator is co-referent with the undergoer and also with the resultee – see the analysis in (14).

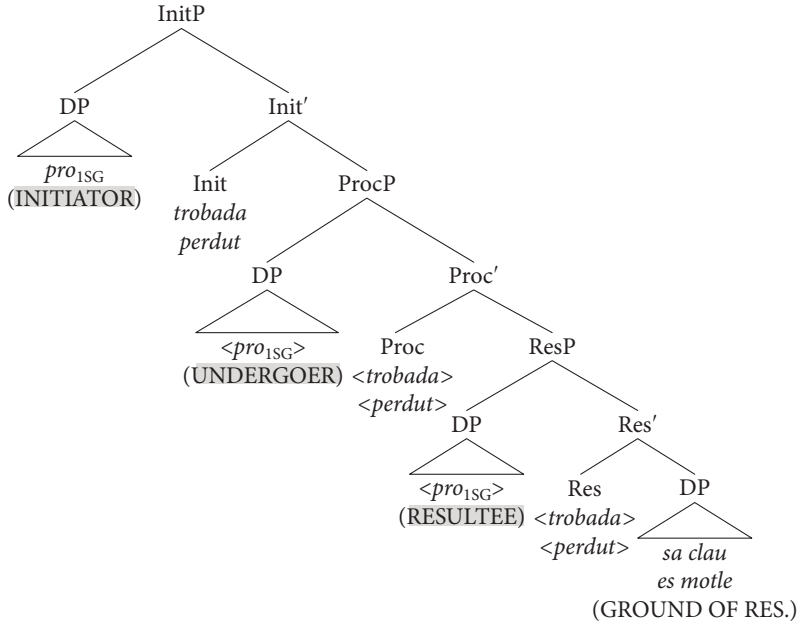
(13) a. *Som trobada sa clau.*
am found.FEM.SG the key
'I've found the key'

b. *Som perdut es motle.*
am lost the mould
'I've lost the mould'

By contrast, with verbs like *menar/conduir* 'drive' and *empènyer* 'push', the initiator is not co-referent with the undergoer (which is the DO). Thus, we can predict that they will not allow, in principle, the auxiliary verb *esser* (15a), unless the initiator could be

also considered as an affected or interested argument (for example, if it maintains a possessive relationship with the DO) (15b). See the two analyses in (16).

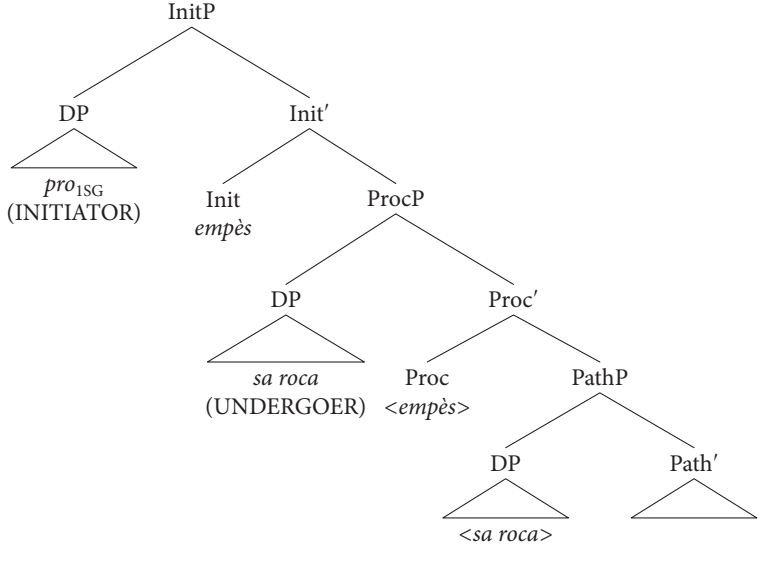
(14)



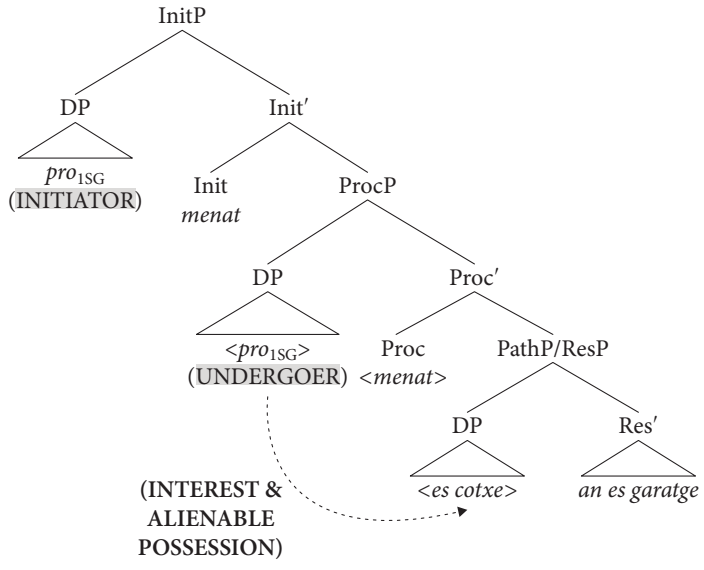
(15) a. *{He/*Som} empès sa roca.*
{have.I/*am} pushed the rock
'I've pushed the rock'

b. ***Som** menat es cotxe an es garatge.*
am driven the car in the garage
'I've driven the car (= my car) into the garage'

(16) a.



b.

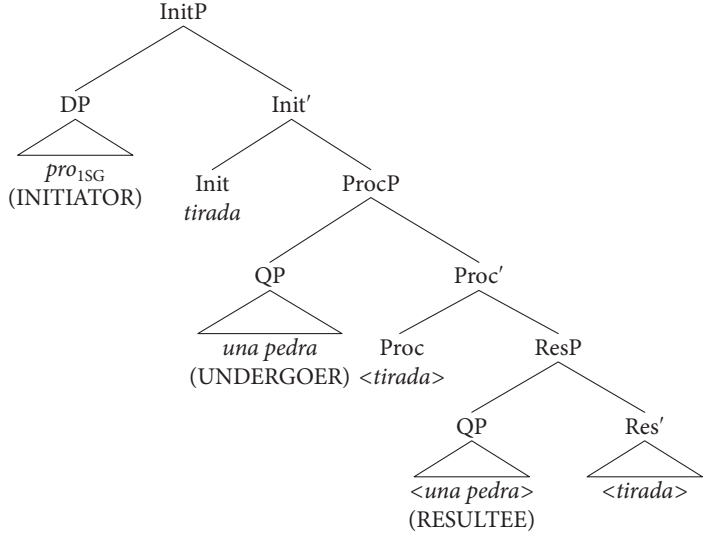


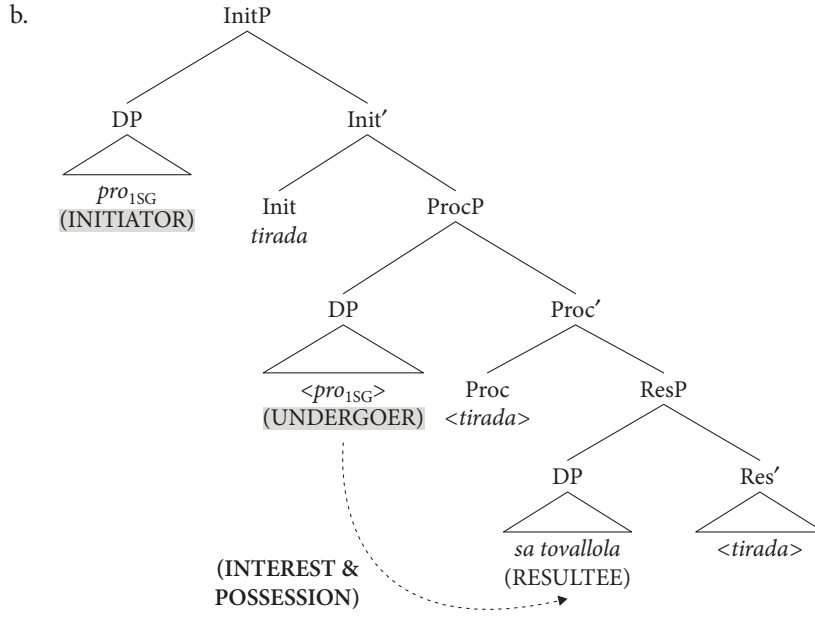
Neither would we expect to find *esser* with verbs like *donar* 'give' or *tirar* 'throw' (17a), nor with causative verbs like *rompre* 'break', *omplir* 'fill'…, which have an initiator and, as a different argument, an undergoer-resultee; unless the initiator could be also conceived as affected or interested (17b):

(17) a. *{He/*Som}* *tirada* *una* *pedra.*
{have.I/*am} thrown.FEM.SG a stone.FEM.SG
'I've thrown a stone'

b. ***Som*** *tirada* *sa* *tovallola.*
am thrown.FEM.SG the towel.FEM.SG
Lit.: 'I've thrown the towel' = 'I've given up'

(18) a.





5. A mixed system of auxiliary selection: person-driven and event-driven

We can conclude that the Old Majorcan system for auxiliary selection is *mixed*: (i) *person-driven* and (ii) *event-driven*.

(i) As for the first factor, according to Loporcaro (2007), where precisely (in which persons) the morphemes 'have' and 'be' occur is a matter of morphology, not syntax. Since a wide range of possibilities exists in the Romance dialects (HHH, EEE, HHE, EEH, HEE, EHH, HEH, EHE), it would be inconceivable to encode this empirical availability directly into structural categories in a one-to-one correspondence. On the contrary, Manzini and Savoia (2007, 2011) claim that a syntactic account of this considerable variation is exactly what is needed.

According to Manzini & Savoia, the "1 & 2 vs. 3" split distinguishes between *discourse-anchored referents* (persons 1 & 2) and *event-anchored referents* (person 3). By contrast, the "1 vs. 2 & 3" split takes into account the *pragmatic salience of speaker reference*. The "1 & 3 vs. 2" split would be a combination of both principles.

The "1 & 2 vs. 3" split has been defined, in more traditional frameworks, by the feature [±participant in the discourse] or [±local] (Benveniste 1966, Harley & Ritter 2002, Legendre 2010). This split has been independently justified in typological linguistic studies. According to Silverstein (1976) and Dixon (1994), 1, 2 & 3 person pronouns are arranged on a hierarchical scale: 1 & 2 are considered

to prototypically confer agentive properties, whereas 3 pronouns are thought to embed inherent information for patient. In a group of ergative languages (like Dyirbal), 1 & 2 pronouns are marked if they function as objects, while 3 person pronouns are marked if functioning as subjects:

(19)

	1 & 2	**3**
agent	-ø	-ŋgu
patient	-na	-ø

(ii) According to Manzini and Savoia (2011), there are systems where the person split interacts with two additional factors (related to event/argument structure), which determine the auxiliary selection: *transitivity/unaccusativity* and *active/middle-passive voice*.[6]

Thus, in Old Majorcan Catalan, besides the "1 & 2 vs. 3" split, as an additional factor, *event structure* (middle-passive voice and unaccusativity) must be taken into account:

a. *Persons 1 & 2* are completely sensitive to event structure: *esser* with the middle-passive clitic *se*, with passives, and with "unaccusative" constructions, in a broad sense (with affected subjects), as pointed out in §§ 3 & 4.
b. *Person 3* is only sensitive to event structure in one sub-factor: *voice – esser* with *se*-constructions and passives, as shown in (1c–d) –, but it is not sensitive to "unaccusativity".

To close this section, we would like to outline the hypothesis that the *generalised person-driven system* found in Northern Catalan (even though 1, 2 & 3 are still sensitive to the middle-passive voice) and in North-Eastern Central Catalan may be the historical next step (or simplification) of a *mixed system* (also *event-driven*) like the one outlined here for Old Majorcan Catalan.[7]

6. For instance, in the dialects of Aliano and Colledimacine, 3SG shows an auxiliary selection according to event structure, while 1SG & 2SG are associated to a single auxiliary verb: either 'have' (in Aliano) or 'be' (in Colledimacine).

7. D'Alessandro & Roberts (2010) provide a minimalist explanation (based on *Agree* in person features) for a *pure person-driven system* like the one found in Eastern Abruzzese. Nevertheless, according to Bonet and Mascaró (2016) and Bonet et al. (2019), synchronically, pure systems do not have split auxiliary selection *stricto sensu*, but just one single auxiliary verb for compound tenses, with allomorphic variation in the root.

6. Final remarks

6.1 A resultative construction?

We have been assuming so far that sentences like (2)–(4) contain a *perfect participle* construction and that we are dealing with the *auxiliary selection* phenomenon. However, Old Majorcan Catalan displays a *sporadic* auxiliary selection, with HAVE/BE optionally alternating. So it could be the case that, in fact, we are not dealing with auxiliary selection *stricto sensu*, but with two different structures and two kinds of participles – as in Larsson's (2015) account for Scandinavian languages –:

a. A construction with a *perfect participle* + HAVE as the auxiliary verb.
b. A construction with a *resultative participle* + BE as the copula.

In such case, we could make the following predictions for the resultative construction: (*a*) the experiential reading and the frequency adjunct '*x* times' would not be felicitous with *esser*; (*b*) manner adverbs (such as *ràpidament* 'quickly') would not be possible, since they are incompatible with states; (*c*) a temporal adjunct introduced by *des de* ('since') would imply that the event described by the participle took place *before* certain moment. However, that is not the case – for instance, in (3f) we have the adjunct *molts pics* ('many times').

Furthermore, the Old Majorcan data do not support the following claim (underlining added): "Notably, auxiliary selection, but not the possibility of BE in resultatives, can depend on tense, mood, person and number, and it does not always correlate with the structure of the VP" (Larsson 2015: 180). Crucially, we still have to explain why *esser* appears, in Old Majorcan Catalan, only with 1 & 2 affected subjects.

6.2 A nanosyntactic approach

Let us approach the Old Majorcan auxiliary selection from another perspective. We will assume the *nanosyntactic* proposal (Starke 2005, 2009; Fábregas 2007, 2009; Ramchand 2008; Caha 2009; *a. o.*) that *lexical insertion comes after syntax.* The primitive building blocks of grammatical computation are features, and the output of the combinatorial operations will target an item from the lexical repertoire. Nanosyntax allows *multi-attachment* or *phrasal Spell-Out:* insertion of exponents under non-terminal nodes – *exponents* are morphophonological

forms of chunks of syntactic structure. Several principles haven been proposed in order to account for the relation between syntax and the set of exponents, such as:

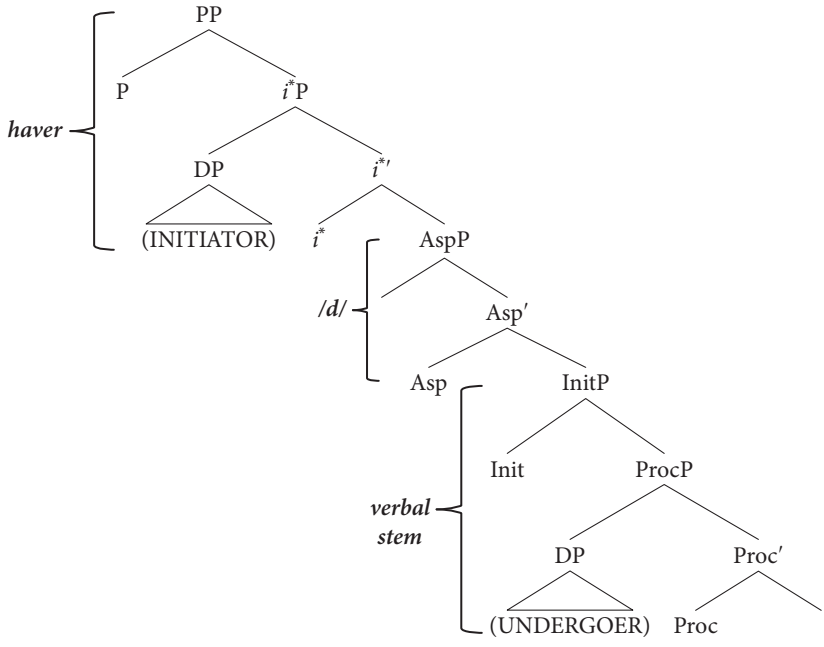
i. *Exhaustive Lexicalization Principle* (Fábregas 2007; Ramchand 2008): "Every syntactic feature must be identified by a lexical item, even if this item is phonologically null".
ii. *Superset Principle* (Starke 2005; Caha 2009): "A vocabulary item matches a node if its lexical entry is specified for a constituent containing that node".

Further assumptions:

1. Freeze (1992) and Kayne's (1993) idea that HAVE is "BE + an incorporated P".
2. Wood & Marantz's (2017) unification of Voice, Appl and *p*. All these heads are, in fact, *i**, a relational head whose specifier allows to introduce an extra argument.
3. Ramchand's (2018) modification of her 2008 model: the *initiator* is no longer introduced as [Spec, Init], but as [Spec, EventP]. Init is just a causative head, with no specifier. Thus, Event could be practically equated to Kratzer (1996) and Pylkkänen's (2008) Voice head (now, *i**).
4. The participle can be derived merging Asp with InitP. The object can move to [Spec, Asp] and, as a result, past participle agreement takes place, with some semantic effects (topicality, specificity).
5. In Old Majorcan Catalan, a Person head can be merged very low in the structure, but it is just active when it possesses a [+participant] or [+local] feature – or a reflexive feature. Independent reasons support the presence or relevance of a Person head in Majorcan Catalan:
    a. The traditional clitics order in Majorcan Catalan is pervasively DO IO: *la me, la te*, where the 1st/2nd clitics would be already licensed from a lower position.
    b. Verbs in 1SG of present take a Ø morpheme: *jo cant* 'I sing', *jo corr* 'I run', *jo dorm* ('I sleep'), *jo dividesc* 'I divide'. The [person] features in T would not matter, since they are already licensed from below.
    c. Prepositions in Majorcan Catalan can take the unmarked nominative case: *a jo* 'to me', *amb jo* 'with me' – by contrast, in Standard Catalan they take oblique case: *a mi, amb mi*.

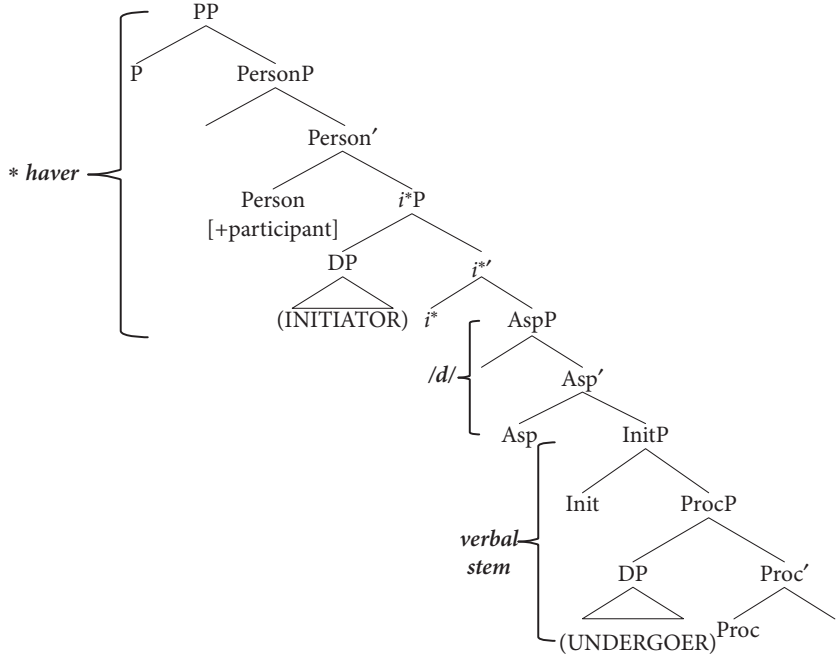
The chunk [P, *i**] would be lexicalized as the auxiliary *haver*, as in (20):

(20)



However, if Person$_{[+participant]}$ is present in the structure – or active – , the chunk [P, i*] is interrupted, so *haver* cannot lexicalize it:

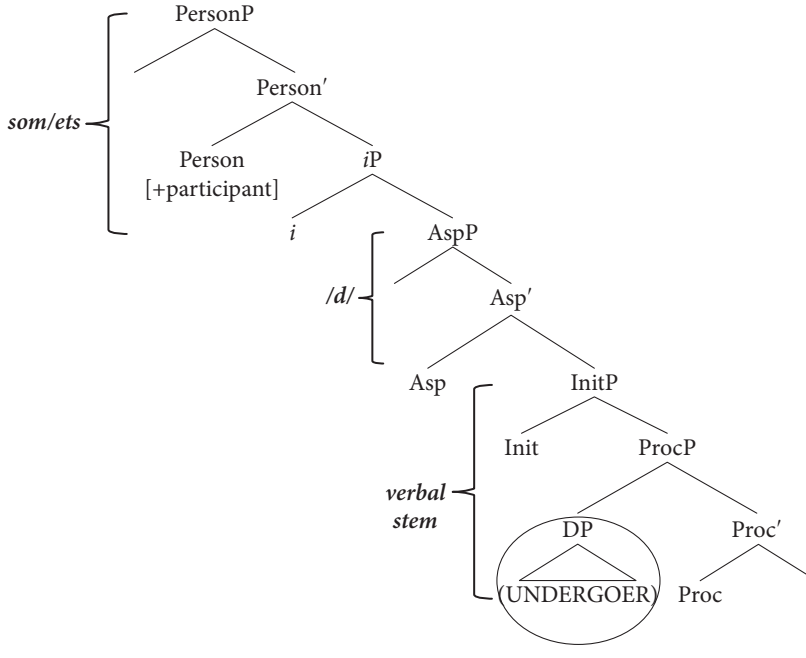
(21)



Now we are left with two possible solutions:

a. Do not Merge the active *i**P projection (or $\text{Voice}_{\text{active}}$P), but an inactive *i*P (or $\text{Voice}_{\text{inactive}}$P), with no specifier. The observed fact that, in these cases, the subject can be interpreted as non-volitional or unintentional (§§ 3 & 4) could be derived from the absence of *i** and from the necessary *undergoer* characterization of the subject (merged in [Spec, ProcP]). However, notice that a causative Init head is still present in the structure.

(22)



b. Or we could build the standard structure in (20), with no PersonP. Therefore, the sporadic/optional status of the split auxiliary selection in Old Majorcan Catalan would be explained, since, in fact, there are two different structures available, with some little semantic differences between them: (20) and (22).

Funding

This research was supported by grants from the Spanish Ministerio de Economía y Competitividad (FFI2014-56968-C4-1-P and FFI2017-87140-C4-1-P) and from the Generalitat de Catalunya (2014 SGR 1013, 2017 SGR 634, FI-DGR 2016 and FI-DGR 2017).

Acknowledgement

I would like to thank Antonio Fábregas, Anna Gavarró, Jaume Mateu, Anna Pineda, Josep Quer, Gillian Ramchand, Gemma Rigau and two anonymous reviewers for their useful advice.

References

Alcover, Antoni M. 1903. "Questions de llengua y literatura catalana." *Bolletí del Diccionari de la Llengua Catalana* 1:15.

Badia i Margarit, Antoni M. 1981. *Gramàtica històrica catalana*. València: Tres i Quatre (Biblioteca d'Estudis i Investigacions).

Badia i Margarit, Antoni M. 1994. *Gramàtica de la llengua catalana: Descriptiva, normativa, diatòpica i diastràtica*. Barcelona: Proa.

Benveniste, Émile. 1966. *Problèmes de linguistique générale*. Paris: Gallimard.

Bentley, Delia. 2016. "Split intransitivity." In *The Oxford Guide to Romance Languages*, ed. by Adam Ledgeway, and Martin Maiden, Chapter 50. Oxford: OUP.

Bonet, Eulàlia, and Joan Mascaró. 2016. "Phase Theory and Prosodic Domains." Bellaterra: CLT seminar (handout).

Bonet, Eulàlia, Lisa L. S. Cheng, Laura Downing, and Joan Mascaró. 2019. "(In)direct Reference in the Phonology-Syntax Interface under Phase Theory: A Response to "Modular PIC" (D'Alessandro and Scheer 2015)". *Linguistic Inquiry* 50(4): 751–777.

Caha, Pavel. 2009. "The Nanosyntax of Case". Ph.D. thesis, University of Tromsø.

Cuervo, M.ª Cristina. 2008. "La alternancia causativa y su interacción con argumentos dativos." *Revista de Lingüística Teórica y Aplicada* 46: 1.

D'Alessandro, Roberta, and Ian Roberts. 2010. "Past Participle Agreement in Abruzzese: Split Auxiliary Selection and the Null-Subject Parameter." *Natural Language and Linguistic Theory* 28.

Dixon, Robert. 1994. *Ergativity*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511611896

Fabra, Pompeu. 1922. *Converses filològiques* (17-12-1922).

Fábregas, Antonio. 2007. "The Exhaustive Lexicalization Principle." *Nordlyd* 34. University of Tromsø, 165–199.

Fábregas, Antonio. 2009. "An Argument for Phrasal Spell-Out: Indefinites and Interrogatives in Spanish." *Nordlyd* 36 (1). University of Tromsø, 129–168.

Freeze, Ray. 1992. "Existentials and other Locatives." *Language* 68: 553–595. https://doi.org/10.2307/415794

Hale, Ken, and Samuel J. Keyser. 2002. *Prolegomenon to a Theory of Argument Structure*. Cambridge (Mass.): MIT Press. https://doi.org/10.7551/mitpress/5634.001.0001

Harley, Heidi, and Elizabeth Ritter. 2002. "Person and Number in Pronouns: A Feature-Geometric Analysis." *Language* 78: 484–526. https://doi.org/10.1353/lan.2002.0158

Jaque, Matías. 2014. "La expresión de la estatividad en español: Niveles de representación y grados de dinamicidad." Ph.D. thesis, UAM, Madrid.

Kayne, Richard S. 1993. "Toward a Modular Theory of Auxliary Selection." *Studia Linguistica* 47: 3–31. https://doi.org/10.1111/j.1467-9582.1993.tb00837.x

Keller, Frank, and Antonella Sorace. 2003. "Gradient Auxiliary Selection and Impersonal Passivization in German: An Experimental Investigation." *Journal of Linguistics* 39, 57–108. https://doi.org/10.1017/S0022226702001676

Kiparsky, Paul. 1998. "Partitive Case and Aspect." In *The Projection of Arguments: Lexical and Compositional Factors*, ed. by Miriam Butt, and Wilhelm Geuder. Stanford (California): CSLI.

Kratzer, Angelika. 1996. "Severing the External Argument from the Verb." In *Phrase Structure and the Lexicon*, ed. by J. Rooryck, and L. Zaring, 109–138. Dordrecht: Kluver. https://doi.org/10.1007/978-94-015-8617-7_5

Larsson, Ida. 2015. "The HAVE/BE Alternation in Scandinavian – Perfects, Resultatives and Unaccusative Structure." In , *Auxiliary Selection Revisited. Gradience and Gradualness*, ed. by R. Kailuweit, and M. Rosemeyer, 145–182. Berlín: Deruyter.

Legendre, Geraldine. 2010. "A Formal Typology of Person-Based Auxiliary Selection in Italo-Romance." In *Syntactic Variation: The Dialects of Italy*, ed. by Roberta D'Alessandro, Adam Ledgeway, and Ian Roberts, 86–101. Cambridge: CUP.

Loporcaro, Michele. 1998. *Sintassi comparata dell'accordo participiale romanzo*. Torino: Rosemberg & Sellier.

Loporcaro, Michele. 2007. "On Triple Auxiliation in Romance." *Linguistics* 45: 173–222. https://doi.org/10.1515/LING.2007.005

Manzini, M. Rita, and Leonardo M. Savoia. 2007. *A Unification of Morphology and Syntax: Studies in Romance and Albanian Dialects*. Londres: Routedge.

Manzini, M. Rita, and Leonardo M. Savoia. 2011. *Grammatical Categories: Variation in Romance Languages*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511974489

Massanell, Mar .2017. Los atlas lingüísticos como fuente para estudios de sintaxis dialectal: El caso del *Atles lingüístic del domini català* (ALDC)." *Verba* 44: 271–315. https://doi.org/10.15304/verba.44.2949

Mateu, Jaume. 2002. "Argument Structure. Relational Construal at She Syntax-Semantics Interface." Ph.D. dissertation, Universitat Autònoma de Barcelona.

Mateu, Jaume. 2009. "Gradience and Auxiliary Selection in Old Catalan and Old Spanish." In *Historical syntax and linguistic theory*, ed. by P. Crisma, and G. Longobardi, 176–193. Oxford: OUP. https://doi.org/10.1093/acprof:oso/9780199560547.003.0011

Moll, Francesc de Borja. 1952 (2006). *Gramàtica històrica catalana*. València: Universitat de València.

Moll, Francesc de Borja. 1975. *Gramàtica catalana referida especialment a les Illes Balears*. Palma: Editorial Moll.

Obrador, Bernat. 1906a. "Del ús dels verbs auxiliars catalans." *Actes del I Congrés Internacional de la Llengua Catalana, tema VII*. Barcelona: Estampa d'En Joaquim Horta.

Obrador, Bernat. 1906b. "Resposta a l'esmena addicional de [PUIG I SAIS (1906)]." *Actes del I Congrés Internacional de la Llengua Catalana*. Barcelona: Estampa d'En Joaquim Horta.

Puig i Sais, Armengol. 1906. "Esmena addicional (no acceptada pel ponent) a la conclusió III [de la ponència de B. Obrador]." *Actes del I Congrés Internacional de la Llengua Catalana*. Barcelona: Estampa d'En Joaquim Horta.

Pylkkänen, Liina. 2008. *Introducing Arguments*. Cambridge (MA): MIT Press. https://doi.org/10.7551/mitpress/9780262162548.001.0001

Ramchand, Gillian. 2008. *Verb Meaning and the Lexicon: A First-Phase Syntax*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511486319

Ramchand, Gillian. 2018. *Situations and Syntactic Structures: Rethinking Auxiliaries and Order in English* [Linguistic Inquiry Monograph 77]. MIT Press, 2018. https://doi.org/10.7551/mitpress/10911.001.0001

Ramos, J. Rafel. 2005. "Els auxiliars dels temps compostos en català medieval." *Caplletra* 38: 179–209.

Rigau, Gemma. 1998. "La variació sintàctica: Uniformitat en la diversitat." *Caplletra* 25.

Rosselló, Joana. 2002. "El SV (I): El Verb i Arguments Verbals." In *Gramàtica del català contemporani*, Vol. 2, ed. by Joan Solà, M. Rosa Lloret, Joan Mascaró, and Manuel Pérez Saldanya, 1853–1949. Barcelona: Empúries.

Salvà, Sebastià. 2015. "Concordança del participi i aspecte intern en el català de Mallorca". final master thesis, màster interuniversitari de Ciència Cognitiva i Llenguatge (CCiL), Barcelona.

Salvà, Sebastià. 2017. "Past Participle Agreement in Majorcan Catalan: The Relevance of Inner Aspect." *Borealis: An International Journal of Hispanic Linguistics* 6:1, 53–75. https://doi.org/10.7557/1.6.1.4101

Salvà, Sebastià. 2018. "La concordance del participi en el català de Mallorca." *Llengua & Literatura* 28: 53–79.

Silverstein, Michael. 1976. "Hierarchy of Features and Ergativity." In *Grammatical Categories in Australian Languages*, ed. by Robert M. W. Dixon, 112–171. Canberra: Australian Institute of Aboriginal Studies.

Sorace, Antonella. 2000. "Gradients in Auxiliary Selection with Intransitive Verbs." *Language* 76: 859–890. https://doi.org/10.2307/417202

Sorace, Antonella. 2004. "Gradience and Implications for Unaccusativity." In *The Unaccusativity Puzzle. Explorations of the Syntax-Lexicon Interface*, ed. by A. Alexiadou, E. Anagnostopoulou, and M. Everaert. Oxford: OUP.

Starke, Michal. 2005. "Nanosyntax." *Seminar taught at CASTL*, University of Tromsø.

Starke, Michal. 2009. "Nanosyntax: A Short Prime to a New Approach to Language." *Norlyd* 36 (1): University of Tromsø, 1–6.

Veny, Joan, and Mar Massanell. 2015. *Dialectologia catalana: Aproximació als parlars catalans*. Barcelona, Alacant, València: Publicacions i Edicions de la Universitat de Barcelona, Publicacions de la Universitat d'Alacant, Publicacions de la Universitat de València.

Wood, Jim, and Alec Marantz. 2017. "The Interpretation of External Arguments." In *The Verbal Domain*, ed. by R. D'alessandro, I. Franco, and Á. J. Gallego, 255–278. Oxford: OUP.

Palenque (Colombia)

African language survivals and their identification

Armin Schwegler
University of California, Irvine

This article examines the revitalization of African language survivals in the former maroon community of Palenque, Colombia. During much of the 20th century, lexical Africanisms in everyday Palenquero creole were scant. Ritual language (*Lumbalú*) and the speech of a few elderly Palenqueros still contained isolated Africanisms, but these had become archaic, nearing extinction.

Profound sociolinguistic changes in Palenque in the 1980-90s introduced sentiments of *negritud* ('black awareness') and appreciation for "all things African", *lexicalia* included. Thus, younger Palenqueros – with the help of some *Lengua* teachers – developed lists of Afro-Palenquero archaisms, rescuing them from oblivion. Various mechanisms spread these *lexicalia "africana"* to the wider speech community, where they are embraced as an index of their maroon and Afro-Colombian identity.

Keywords: Africanisms, language and identity, Palenque, Palenquero

1. Introduction

This article examines how recent changes in intergenerational attitudes towards the Afro-Hispanic creole spoken in the former maroon community of Palenque (Colombia) have profoundly influenced the retention and promotion of sub-Saharan Africanisms, thereby fundamentally reshaping speakers' perception, appreciation, and projection of black identity.

As this study hopes to show, these changes in Palenqueros' language attitudes began to gain momentum just prior to this millennium, and now significantly complicate linguists' analysis of the true nature and historic trajectory of African language survivals in this once remote part of the New World. The result of over 30 years of intermittent fieldwork in Palenque (cp. Schwegler 2017a, and relevant sources cited therein), this paper constitutes a follow-up to Schwegler's (1994) publication, and aims to elucidate how and why especially younger Palenqueros

https://doi.org/10.1075/rllt.16.15sch

now routinely hyperpromote words and grammatical constructions of presumed African origins so as to accentuate the community's sub-Saharan ancestry.

After a brief Introduction, Section 2 describes Palenque and its unique sociolinguistic position within black Latin America. Section 3 revisits Palenque's language dynamics during the period of 1950-2000, when the hypervaluation of perceived lexical Africanisms eventually triggered a marked increase in the use of sub-Saharan vocabulary. Section 4, the core of this study, updates my prior examination of this topic by concentrating on intergenerational changes that have profoundly altered language attitudes in Palenque since the start of the new millennium. To properly contextualize this information in time and space, I will examine several aspects of Palenquero language, linking them with the contemporary rise of sentiments of *negritud* (black awareness) in Colombia at large. Section 5 examines the wider significance of the findings, and offers a summary and conclusion.

2. Palenque: Africa in Latin America

By all accounts, Palenque (approximately 4000 inhabitants) is the "blackest" and ethnically most uniform village in Latin America, a fact confirmed by recent population genetics research (Martínez et al. 2017, Noguera et al. 2014). DNA gathered in Palenque shows that approximately 75%-80% of their genetic roots are sub-Saharan, the remainder being mostly European (American Indian contributions are negligible).

Located some 60 kilometers from Cartagena de Indias (Map 1) – once the "blackest" and most ethnically diverse city in the New World (Schwegler 2014: 408-409) – Palenque has always been considered a "case sui generis". The Palenqueros have, for instance, consistently insisted that after their escape from slavery into the lush interior of the State of Bolívar, their *raza* ('race, ethnic group') shunned mixing with outsiders, thus remaining "profoundly African," genetically as well as socioculturally (Ansari Pour 2010, Ansari Pour et al. 2016, Martínez et al. 2017, Noguera et al. 2014, Schwegler 2006, 2007, 2016b). Over the years, their abject poverty, lack of access to schooling and/or modern amenities, and especially their skin color – visibly darker than that of the surrounding Afro-Hispanic communities – were all factors that contributed significantly to the extraordinary stigma that outsiders attached to Palenque's population and culture (creole language included).[1]

1. Within the context of Cartagena de Indias' modernization and globalization (it now has a booming international tourist industry), Palenque's perceived backwardness is in many ways

Map 1. Location of Palenque

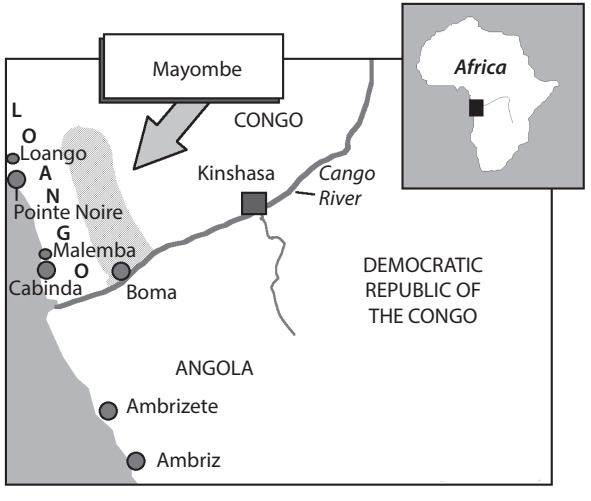
Throughout much of the 20th century, this persistent social stigmatization and marginalization exerted enormous pressure on the community. As a result, prior to the new millennium, lexical Africanisms (Schwegler 2000, 2002a) came to be viewed within the village as *anticuados* ('outdated') and thoroughly detrimental to

real. Electricity, for instance, was still absent or at best intermittent in most households almost to the year 2000, and a sewer system and running water were non-existent prior to 2015.

then emerging ideas of *progreso* ('progress, social advancement, modernization'). In reaction to these social pressures, starting around 1970, Palenque experienced rapid generational language abandonment, away from the creole and in the direction of Spanish, spoken natively in the village for over three centuries. This shift accelerated during the 1980s, so much so that by the mid-1990s, Spanish monolingualism had virtually become the norm among the young (most, if not all, Palenqueros did, however, still have passive knowledge of the creole). Prevailing local and regional attitudes toward Palenquero were so negative that neither locals nor linguists had any reason to believe that the creole would survive beyond the first half of the 21st century.

Notwithstanding these momentous sociolinguistic shifts within the village, any outsider fortunate enough to reside among the Palenqueros for extended periods from 1975 to 2000 would have noticed the unequivocally non-Western nature of several of their traditions, several of which have been reported in Schwegler (1994, 1996, 2006, 2007).

Researchers first expected Palenque to have a diverse African substrate (Kwa, Bantu, etc.). These expectations were, however, ultimately met with an entirely contrary finding: thanks to rapid advances in collaborative scholarship involving historians, anthropologists and especially linguists and population geneticists, we now know that Kikongo – spoken by the Bakongo of western Central Africa – was Palenque's only significant African substrate, and that the small region of Mayombe (Map 2) must have been the center of the Palenqueros' original habitat (Schwegler 2017a: 108; see also Moñino forthcoming and Schwegler 2016b).



Map 2. Mayombe region, likely homeland of many of Palenque's original maroons

Over the years (1990–2010), in Palenque two locally-born school teachers got wind of these scholarly findings, and began "spreading the origins news" in their classrooms and beyond. This and other related events in turn triggered a community-wide retelling of Palenque's heroic past, elevating the question of "African origins" to levels not seen before in Palenque.

Palenqueros' enthrallment with their African roots persists to this day. Almost twenty years into the new millennium, Palenqueros now routinely promote "origin ideas" that invoke their heroic "Bantu past", "Congo" (or "Mayombe") origins, and direct blood link to the heroic rebel slave Benko Bioho, whose imposing statue adorns Palenque's Central Plaza.[2] Often, this recasting of Palenquero history takes place in front of academic tourists (Colombian high school teachers and their pupils, national and international scholars including anthropologists, linguists, musicians, etc.), who flock to Palenque on an almost daily basis. There, locals regularly tell them that "Palenque es un pequeño rincón de África en América" ('Palenque is a little corner of Africa in the Americas').

As shown in Schwegler (2017a), a good portion of these collective memories serve to strengthen locals' appreciation for *negritud* ('black awareness, black pride'), and to highlight the important role that remaining linguistic Africanisms currently play within the context of Palenquero identity (re)building (Moñino 2003). As the following sections hope to reveal, this recasting of history and language presents fresh research opportunities as well as a potential minefield for scholars interested in understanding and documenting the true trajectory of Palenque's African lexical survivals.

3. Language dynamics from 1950–2000: Hypervaluation of perceived lexical Africanisms

Subjected to the continuous cultural, linguistic, and socioeconomic domination of the nearby metropolis of Cartagena (over 1 million inhabitants today), 20th-century Palenqueros felt an ever-greater need to justify and define their peculiar culture and ways of speaking. While they used multiple strategies to counteract the perceived outside threat, one response in particular came to be favored: self-identification and promotion of presumed lexical Africanisms that, mostly on the basis of (a) phonetic features and (b) contrastive analysis with Spanish, had

2. For photos of the statue and an in-depth essay of its role and meaning in Palenque, see Ferrari (2017).

a distinctly sub-Saharan flavor. The underlying mechanics behind this contrast was simple: if a lexical item in question was absent from Spanish, it was generally considered a strong candidate for inclusion in Palenqueros' list of Africanisms. Anxious to reaffirm their African roots, Palenqueros began to attribute sub-Saharan origins to creole words whose derivation in many cases was not African but instead European (Schwegler 1994, 2002a).

To properly contextualize Palenque's sociolinguistic dynamics of the 1970s and beyond, mention must be made of two factors that contributed significantly to the hypervaluation of presumed Africanisms. The first concerns the relative paucity of lexical Africanisms that at the time survived in the village. Upon their initial contact with the Palenquero language, even seasoned creolists such as Germán de Granda and Derek Bickerton were surprised by the exceedingly low number of lexical Africanisms in daily Palenquero discourse, numbering, as they did then, a mere dozen or so (Granda 1968; Bickerton & Escalante 1970). Most prominent among these were:[3]

ngombe	'cow, steer, cattle,'	*moná*	'youngster, kid'
lumbalú	'funeral chant'	*mahaná*	'children, youngsters'
kasimba	'water hole'	*musá*	'cornmeal (dish)'
tusí	'excrements, shit'		

This dearth of Africanisms in daily creole parlance explains, in part, why Palenqueros ultimately valued a select few words of European origin with a distinctively African flavor as especially precious remnants of their sub-Saharan past. Subsequent attention by scholars to Africanisms eventually did reveal 200–300 additional Palenquero words of genuine African origin, but these were almost always archaic, virtually never used, and, up to the 1990s, generally known only to the eldest in the village.

The second factor that led to an exaggeration of Palenque's African linguistic legacy was originally external to Palenque's community, and rooted in well-meaning attempts by early scholarly publications to etymologize "exotic" Palenquero vocabulary. The fact that well over 90% of all Palenquero lexicon could readily be linked to Iberian sources led scholars to assume that the remaining tokens *a fortiori* had to have African roots. Not unexpectedly, perhaps, this *modus operandi* led to a series of mistaken etymological hypotheses that eventually reached Palenque in the form of published pamphlets (photocopies, offprints, etc.), from where they emanated to a select few but influential members of the community.

3. See Schwegler (2002a).

To be sure, not all attempts by Palenquero *aficionados* of African words turned out to be ill-fated. Especially starting in the late 1980s and early 1990s, several genuine Africanisms did receive special attention and community-wide promotion, thereby revitalizing African words once virtually doomed to disappear. This included expressions such as *mulumba* 'assailant with sexual intent' (cp. also *mulumbiá* 'to chase after women [with sexual intent]') and *kutú* 'force, strength', both of which are firmly entrenched in today's Palenquero lexicon, and are appreciated as expressions of both the truly local and genuinely "African". As might be expected, no Palenquero alive today could relate *mulumba* and *kutú* to their respective Kikongo etyma (e.g., *mu+ndúmba* 'young woman, girl, virgin, Miss, etc.', Schwegler 2002a: 204, Laman 1964: 674). However, the combination of their phonetic configuration and their non-existence in Spanish are generally viewed as sufficient evidence by the Palenquero community to hold them in special esteem, and consider their legacy as exclusively African.

The sociolinguistic changes in attitudes that gradually transformed life in Palenque from the 1970s to the early 1990s were followed by two momentous events that, at the time, seemed rather insignificant. These took place in a context where the village and its inhabitants still suffered from the earlier-mentioned deeply negative sentiments in reference to black culture at large, and to black language in particular (Schwegler & Morton 2003: 101–111). As noted in Schwegler (2011), by the 1990s, Palenqueros had become convinced that the autonomous *Lengua* had entered the final stage of a lengthy journey that had begun with their forefathers' heroic escape from slavery. Many felt that an accelerated transition to Spanish monolingualism would facilitate Palenque's integration into the modern world. Not surprisingly, even older Palenqueros began to avoid using *Lengua* altogether, so much so that, at least on the surface, Palenque began to function as a quasi-monolingual society.

Among Palenque's youths, there were, however, a few who dared to buck these trends, thereby laying the foundation for the astounding revitalization of Palenque's *Lengua* and African heritage that would gain true momentum some ten to fifteen years later (Pfleiderer 1998; Morton 2005; Schwegler 2011; Lipski 2012). Part of this revitalization included activities that lay the groundwork for compilations (by Palenquero *Lengua* teachers and their students) of "African" vocabulary lists, obtained by way of enthusiastic local "word hunts" whose goal was to rescue highly antiquated African vocabulary. Some of these archaisms subsequently experienced an almost miraculous revival, as they have now become very frequent expressions of Afro-Palenquero identity and pride. Among them is the formerly moribund *enu* 'you (PL.)' < Kikongo *énu* 'you [PL., emphatic]'), which by 2009 had made its way from scholarly works (e.g., Friedemann & Patiño 1983; Schwegler 2002b) into Palenque's classrooms, and from there into the barrios of the village.

All the while, the generational language abandonment that had gripped Palenque for over a quarter century continued unabated, and there were no indications yet that Palenquero speech conventions would some day be supplanted by a community-wide fervor for the creole and the Afro-Hispanic local culture for which it stands.

4. Palenque in the new millennium: *Lengua* revitalization, ethnic pride and the community-wide reappraisal of African archaisms

4.1 Sociolinguistic changes post 2000

Details surrounding the extraordinary post-2000 revival of *Lengua* are given in Schwegler (2011; see also Lipski 2012 and Moñino 2017). The following paragraphs offer a brief sketch of the mechanisms by which sociolinguistic engineering has recently reshaped life in Palenque, thereby leading to its full(er) integration into Cartagena's regional sphere.

Prior to the 1990s, in Colombian society and academia there prevailed a widespread disinterest and apathy regarding the promotion or study of black culture. So widespread and deeply engrained were these negative sentiments in reference to black culture (language included) that the Palenqueros themselves had largely embraced these same values, and, partially as a result of these attitudes, passively accepted the rapid decline in the use of *Lengua*.

By the mid-1990s, Palenque had begun to attract significant attention in international academic circles resulting in a relatively steady presence of visitors in the village. Most of these outsiders were intent on witnessing for themselves how Palenqueros had managed to preserve an exotic "black" vernacular within the larger context of modern Colombian society. As the new millennium approached, this recognition given to Palenqueros by outsiders awoke a sense of pride among a few Palenqueros, leading them to also publicly reject – for the first time – the stigmatization that had always attached to their culture (and *Lengua*). Black pride and ethnolinguistic awareness, it seemed, were beginning to make themselves felt in earnest for the first time in Cartagena's hinterland. These small-scale local measures eventually coincided with Colombia's broad-based national movement of *negritud* ('black awareness'), also felt in other parts of Latin America (e.g., Peru, Ecuador, and Venezuela).

Colombia's promotion at the close of the 20th century of linguistic and cultural recognition of ethnic minorities created an air of profound national discourse centered upon race, culture and history (Schwegler & Correa 2018). This discourse has continued unabated into the second decade of the new millennium, all of it significantly influenced by the recasting of the 1991 Constitution, which

liberalized social and economic policies, and granted fundamental rights to ethnic minorities (Afro-Colombians included).[4] The same document recognized and prominently situated Palenque's vernacular at the center of its black patrimonial heritage. Inspired and encouraged by these and related sociopolitical events *outside* their community, some Palenquero youths developed a rapidly growing sense of empowerment, and a willingness to defend linguistic and cultural rights that they were previously denied. *Lengua* soon evolved into the ultimate symbol of local pride, ethnic celebration and *negritud.* And within *Lengua,* perceived Africanisms were now openly appreciated as primary bearers of ancestral black roots, deeply anchored in a context of heroic maroonage.

Between 2000 and 2019, signs of Palenque's rapid integration into the modern world ("globalization") and contemporaneous admiration as a sociocultural island onto itself became plentiful, and plainly visible for all. Over the past quarter century, these events have increased Palenqueros' language loyalties, trending away from the once all-powerful Spanish in the direction of now high prestige Palenquero – a state of affairs unimaginable just a few years ago (Schwegler 2011).

By around 2010, interest in Palenquero language and culture had reached feverish levels among some, so much so that they dared to write creole language on the walls of buildings, tomb stones, and modest shops – for all to see and appreciate (for a sampling of these, see the many photos in Schwegler, Kirschen & Maglia 2017). In these public displays, "African-sounding" words were often given preference to underscore the *negritud* (blackness) of Palenque (see, for instance, Photo 3a in Schwegler 2017a: 85, where Kikongo-derived *charamuka* is used as an overt expression of local pride in the Palenque's unique heritage).

Today, the community at large is genuinely proud of its *Lengua,* and generally welcomes the many changes that have transformed their village. Although the long-term survival of the creole is far from assured and even deemed slim by some (cp. Moñino 2017: 15-16), all Palenqueros recognize that it *has* been given a new life-line.

4.2 The community-wide reappraisal of lexical African archaisms

Around the year 2000, the earlier-noted fascination for Africanisms by a few Palenquero youth experienced a momentous shift on two levels: first, the process of revitalization became institutionalized through its integration into Palenque's

4. In this context, see the "Decenio afrocolombiano (2015–2024)", promoted and explained at these official websites by Colombia's Government:
http://dacn.mininterior.gov.co/decenio-afro/que-es-decenio-afro
http://www.un.org/es/events/africandescentdecade/background.shtml

schools, and second, the penchant for perceived Africanisms and concomitant allegiance to *negritud* began to alter creole language use beyond the strictly lexical domain, thus subtly altering creole grammar to a point where it began to generate a "new kind of Palenquero" (cp. Lipski 2012).

4.2.1 *Integration of* Lengua *lessons into the school curriculum*

During the last decade or so, some local teachers have begun to impart *Lengua* lessons in Palenque's schools. Modest at first, this language curriculum has become an integral part of a wider and growing platform of *Etnoeducación*, which favors a hands-on practical approach that engages students with the community at large. One such engagement takes place in the form of the earlier-mentioned "word hunts" (interviews of elders), which has gradually led to compilations of several hundred words of presumed African origin (e.g., see Schwegler 2017a: 87). These lists of indigenous words typically consisted of a mix of Spanish- and African-derived items. A good portion of them have demonstrable sub-Saharan origins; others are unquestionably misidentified as *Afro-Palenquerismos*.

The net effect of this newly obtained *collectanea* of Afro-Palenquero words is that teachers and students routinely share these lists and especially the most African-sounding words contained therein with fellow pupils, and, to a lesser extent, the community at large. In this manner, exotic-sounding archaisms like *kankamaná* 'leader, boss, chief' (< Kik. *ka+ `nka + ma 'áana* lit. 'the leader/chief of the young(sters)' Schwegler 2002a: 192) have made their way into daily Palenquero parlance, acquiring special ethnolinguistic significance in the process. In the case of *kankamaná,* for instance, the term quickly spread outside the immediate sphere of the *Lengua* scene, and came to be favored in official local documents, names of youth groups, as well as in book titles (cp. *Reglamento interno para la administración territorial del consejo comunitario "Ma Kankamaná" de San Basilio* (Martínez Miranda et al. 2014, my emphasis). Today, "Ma Kankamana" serves as name of Palenque's *consejo comunitario* ('Community Council') – the community's "máxima autoridad administrativa del territorio ancestral colectivo[5]" (Martínez Miranda et al. 2014: 7).

This reprocessing of Palenque's African legacy gained momentum once Palenquero aficionados themselves (rather than "just" scholars from the outside) began to publish grammars and dictionaries in the post-2000 period. These publications had a decidedly applied slant, and were always intended to further heritage *Lengua* teaching and reaffirm ethnic pride. Especially noteworthy among these publications are:

5. Lit. 'the highest administrative authority of the ancestral collective territory'.

Diccionario. Lengua afro palenquera-español [= lit. *Dictionary. Afro-Palenquero Language*] by Solmery Cásseres Estrada, 2005. Republished in 2010 (2d edition) under the title *Diccionario AfroPalenque – Español / Español – AfroPalenque* [sic]

Lengua ri Palenge. Jende suto ta chitiá, by Rutsely Simarra Obeso, Regina Miranda Reyes & Juana Pabla Pérez Tejedor, 2008.

Chitieno lengua ku ma kuendo [= lit. *Let's speak Lengua through short stories*] by Bernardino Pérez Miranda 2011.

Gramátika ri luénga Palénque: Pa ma lo ke tan komensá [sic] [lit. *Grammar of Palenque's "Lengua": for those who are going to begin,* i.e., *"Lengua" grammar for beginners*], by Luis Simarra Reyes & Álvaro Enrique Triviño Doval, 2007/2012.

Appreciating these language-external circumstances is key to understanding how and why words like *chitiá*[6] 'to speak, talk, converse, etc.' came into vogue, effectively replacing *ablá* 'idem' (< Span. *hablar*) in *Lengua* schoolbooks and locally-produced Palenquero short stories. Starting around 2015, Palenque's youth also began to consistently favor *chitía* over Spanish-derived *ablá* in (cell phone) text messaging. This practice stands in sharp contrast to interactions among similar age groups of 30 years ago, when *chitía* was virtually non-existent. In 1987, the term became resurrected one afternoon during an extended conversation held in Victor Simarra's backyard (see Photo 3 in Schwegler 2017a: 186), where a few *Lengua* enthusiasts and I often gathered to *chitiá en lengua* ('shoot the breeze in Lengua'). From this small inner circle, highly sporadic *chitiá* eventually spread to beyond Victor's immediate barrio, made possible in part by the exceptionally close-knit nature of a relatively small speech community where everyone pretty much knows everyone else.

The above-cited language sources too have contributed significantly to the wide(r) acceptance of reprocessed "African" vocabulary. This is especially true with regards to the younger generations, who, as noted by Moñino (2017) and others, have acquired *Lengua* in school rather in the home, as well as with the help of publications such as the *Diccionario. Lengua afro palenquera-español.* As even a brief perusal of this and similar pedagogical resources reveals, Hispanisms once firmly entrenched in the creole are simply omitted and substituted with

6. *Chitiá* has never been etymologized, but its origin now seems clear: derived from Pal. *chito* ['ʧsito] 'a (small) piece/portion (of food, etc.)' < Kik. *ki-itu* [kiitu] ~ [ʧsiitu] 'part, piece, lot, portion' (Laman 1964: 294; Schwegler 2017a: 88–89), *chitía* literally meant 'to talk/say/speak little or unimportant things', hence 'to chat', and by subsequent post-2000 semantic expansion, simply 'to talk/speak/converse'.

indigenous terms with an "Afro" tang. True to this practice, the *Diccionario* thus altogether omits Spanish-based *ablá,* listing *chitiá* 'to dialogue' instead (Cásseres Estrada 2005/2010: 54).[7] And, to give just one more example, the same dictionary (p. 93) features *lungá* 'to die', to the detriment of *morí* (common in *Lengua*), left out altogether (see also p. 170, MORIR = *lungá*).

Today, the proliferation of this self-identified ancestral lexis continues unabated. From a strictly synchronic perspective, this has led to a cross-linguistically unusual if not downright baffling situation in which the vernacular of the youngest generations generally uses far more Africanisms than that of their parents (many of whom lack fluency in Palenquero), and, to a lesser degree, grand and great-grand parents. Figure 1 conceptualizes this situation, which in diachronic terms resembles an hourglass in which isolated ancestral words emanate from a few elder individuals to youth at large via the mediation of *Lengua* teachers, didactic materials, vocabulary lists, and so forth. From these youth groups the revitalized words are then diffused – more or less gradually so – to the rest of the speech community.

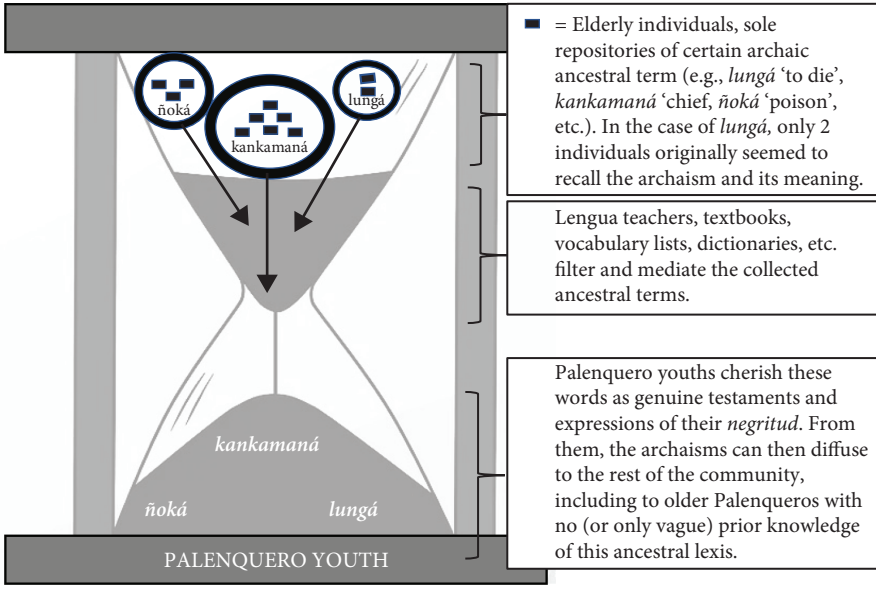


Figure 1. Conceptualization of revitalization process of ancestral vocabulary[8]

7. See also p. 165, where the *Diccionario* similarly favors "exotic" indigenous vocabulary (HABLAR: *chachariá.* 2. *Mambo*) over once ubiquitous Pal. *ablá.*

8. In reality, the transmission of archaisms from the elders to the young is far more complex than can be shown here. In some instances, a word form but not its true meaning is transmitted. In others, revitalized words quickly undergo semantic reinterpretation. This has been

4.2.2 *Grammar prescriptivism and the recasting of creole morphosyntax*

As explained above, for most of its existence, Palenque was a strictly oral society, in which prescriptivism and notions of "good and bad" or "correct and incorrect" *Lengua* were, for the most part, non-existent. With the growing sentiments of *negritud* and the concomitant institutionalization of the creole language curriculum into Palenque's classrooms, this former situation changed rather dramatically. Locally-born educators like Bernardino Pérez Miranda published their first creole grammars (see also Simarra Reyes & Triviño Doval, 2007/2012), which sought to describe creole morphosyntactic constructions in prescriptive ways that maximally distinguished them from Spanish patterns. Palenquero grammar and sentiments of local identity thus became intertwined more deeply, and, as a consequence, prescriptivist teachers and their pupils began to speak a "new Palenquero" (cp. Lipski 2012) – one that differs in subtle ways from that of older generations. Features of "novel and prescriptive grammar" that promote indigenous over Spanish-like constructions can be cited from various domains, including those of nominal pluralization (greater use of plural marker *ma*), pronoun usage (preferential treatment of 1st person pronoun *i* rather than *yo*), and predicate negation. The following examples of predicate negation (cf. Schwegler 1988, 1991, 2016a, 2017a, 2018a, 2018b) may suffice to illustrate the point.

Traditionally (pre-2000), spoken Palenquero discourse has had three different predicate negation patterns (preverbal NEG_1, double NEG_2, and postverbal NEG_3).

(1)	a.	*Ele*	***nu***	*ten*	*monasito.*			NEG_1
	b.	*Ele*	***nu***	*ten*	*monasito*	***nu.***		NEG_2
	c.	*Ele*	–	*ten*	*monasito*	***nu.***		NEG_3
		s/he	NEG	have	child	NEG		
		'He/she doesn't have a child.'						

(2)	a.	*Suto*	***nu***	*kelé-ba*	*kumé*	*kumina.*		NEG_1
	b.	*Suto*	***nu***	*kelé-ba*	*kumé*	*kumina*	***nu.***	NEG_2
	c.	*Suto*	--	*kelé-ba*	*kumé*	*kumina*	***nu.***	NEG_3
		we	NEG	want-TMA	eat	food	NEG	
		'We didn't want to eat the meal.'						

Complex discourse-pragmatic and information-structural principles govern the selection of these three patterns (Schwegler 2016a; Schwegler 2018a). For instance, NEG_1 is generally reserved for matter-of-fact statements that are

---

the case with kankamaná, which formerly denoted "leader of a ritual group, possessor of ritual "secrets and know-how"; today, the term has a wider meaning, denoting simply "leader, boss, chief" in general (number in Palenquero is unmarked, whence *kankamaná* can also have the plural or collective reading "leaders, etc.", which explains the above mentioned *kankamaná* 'community council').

presuppositionally neutral. In contrast, NEG_2 and NEG_3 constructions are presuppositionally loaded, and can occur only with inferable or explicitly activated propositions. NEG_2 and NEG_3 further differ pragmatically in that they each convey differential scalar TRUTH: NEG_3 resets the truth, NEG_2 merely rejects a prior truth assumption.

What matters most in the context of this study is this: courses taught in Palenque's *etnoeducación* program and the texts (grammars) used therein falsely characterize the preverbal NEG_1 pattern as a synchronic intrusion from (regional) Spanish, where NEG_1 is the only predication negation pattern (cp. Span. *yo NO quiero* 'I don't want to'). Consequently, *Lengua* teachers instruct students to avoid the Spanish "intruder" NEG_1, and to use NEG_2 or NEG_3 instead. As a result, younger Palenqueros L2 speakers tend to systematically shun constructions such as *yo NU kelé* (NEG_1). In the process, they forgo the subtle but communicatively important pragmatic nuances available in traditional *Lengua*, and as a consequence, "new" and "old" Palenquero drift apart.

Interestingly, consciously or subconsciously, younger Palenqueros (and their teachers) tend to consider the non-Spanish negation pattern as part of their "Afro" heritage, thereby not only assigning it special status but also intrinsically linking it to sub-Saharan roots. Scholarship has not yet established with certainty whether non-hispanic $\text{NEG}_{2/3}$ are morphosyntactic calques from Kikongo (and/or possibly other African languages) – a hypothesis that has considerable appeal for reasons explained briefly in Schwegler (2017a: 40, fn. 22). One thing is certain though: Palenqueros' sentiments of *negritud* and their inclination to consider all cultural elements (*Lengua* included) with a distinctive non-Hispanic flavor as "African" is the unmistakable engine that currently shapes the re-engineering of the "new Palenquero".

5. Summary and conclusion

As this article hopes to have shown, over the last quarter century, Palenqueros have reprocessed their cultural heritage by purposefully reactivating ancestral words of African origins. This auto-identification of archaic vocabulary is not always "on target", as they also recuperate words like *suebbesuebbe, ablakablaka, (a)ngalá,* and so forth, whose roots are demonstrably European rather than sub-Saharan. In Palenque's local context, this hardly matters, however, as all of these presumed "African" words play an equally important role in (linguistic) community building and feelings of *negritud.* Without question, a good proportion of the revitalized words have become reintegrated into the creole (and, at times, also into Spanish)

not out of communicative necessity but as a supplement to Palenqueros' linguistic repertoire, indexing their maroon history and "Afro" identity. This social dynamic is eloquently captured by Palenquera *Lengua* aficionada Rutsely Simarra Obseso, who expressed her sentiments as follows:

> *Hablar de lengua palenquera […] no es circunscribirse únicamente a los aspectos gramaticales y estructurales que la constituyen, sino pensar en un conjunto de elementos solidarios que recogen las maneras de ser, sentir e interpretar las realidades propias de una comunidad que aún mantiene sus legados ancestrales …*
>
> (Simarra Obseso 2006: 80–91)

> When talking about the Palenquero language, one cannot simply circumscribe oneself to its grammatical and structural aspects; rather, one must view *Lengua* as an amalgam of elements of solidarity that express our ways of being, our ways of feeling and interpreting the very reality in which our community lives and continues to maintain its [African] ancestral legacy …
>
> (Simarra Obseso 2006: 80–91, my translation)

Palenque's evolutionary path toward language revival and regional integration has unfolded in minute incremental steps. These steps were first taken by very few adolescent individuals, several of whom later became community leaders and/or teachers of *Lengua*. Through their efforts, century-long and profoundly deep-seated stigmatization that attached to "anything and everything Palenquero" could gradually be shed, eventually culminating in the contemporary situation where the entire village has gone "from shame to fame" (Schwegler 2017b).

Robust sentiments of local identity, the program of *etnoeducación*, rapid integration into the globalized world, Colombia's post-1990 *negritud* movement, the modernization of Palenque and its ever more close-knit connection to Cartagena and surrounding towns thanks to a newly paved road, are all factors that, in one way or another, have made possible the unexpected revival of Afro-Palenquero words.

For future linguists interested in studying New World African legacies, the external language-changes that have transformed Palenque present special challenges. This is so because, as shown in this paper, the transmission of many of Palenque's Africanisms has been anything but linear: in sharp contrast to *moná, ngombe, mahaná,* words like *kankamaná* or grammatical elements like *enu* 'you (PL.)' – now widely used in the village – were not handed down from generation to generation in daily parlance, but instead remained relegated to ritual speech and/or the parlace of very few elders. Twentieth-century scholarship then documented these and similar terms, and local teachers (aficionados of *Lengua*) and community activists reintroduced them to their community. Other words like *chitiá* had

an analogous trajectory, but differed in that they formerly also circulated (in the mouths of relatively few) outside of the strictly ritual domain, always indexing in-group identity and solidarity.

For all practical purposes, Palenquero creole today is quite rich in African lexicon, which would seem a natural result of its history of maroonage, long-term geographic and social isolation, and so forth. But, as has been shown here, this superficial synchronic perspective is deceiving, as Palenque's sub-Saharan linguistic legacy is largely the result of a more or less artificial (but very real) re-Africanization that has its roots in the 1980s and 1990s.

Acknowledgment

I thank Miguel Gutiérrez Maté and an anonymous reviewer for their helpful comments to an earlier version of this paper.

References

Ansari Pour, N. 2010. "Human Genetic Variation, Relationships of Peoples of Sub-Saharan Africa and Implications for Healthcare." Doctoral dissertation, University College London. (Slight revisions made in 2011).

Ansari Pour, N., Y. Moñino, and N. Bradman 2016. "Palenque de San Basilio in Colombia: Genetic Data Supports an Oral History of a Paternal Ancestry in Congo." *Proceedings of the Royal Society B* (no Vol. number), 1–11. https://doi.org/10.1098/rspb.2015.2980

Bickerton, D., and A. Escalante 1970. "Palenquero: A Spanish Based Creole of Northern Colombia." *Lingua* 24: 254–267.

Cásseres Estrada, S. 2010 (2005). *Diccionario AfroPalenque – Español / Español – AfroPalenque* [sic]. Republished in 2010 (2d edition) under the title *Diccionario. Afropalenque–Español / Español– Afropalenque*. Cartagena: Ediciones Pluma de Mompox S. A.

Ferrari, L. 2017. "San Basilio de Palenque (Colombia). A *performance* of freedom," In Schwegler, Kirschen & Maglia (eds.), pp. 121–145.

de Friedemann, N. S., and C. Patiño Rosselli 1983. *Lengua y sociedad en el Palenque de San Basilio*. Bogotá: Instituto Caro y Cuervo.

Granda, G. de. 1968. "La tipología "criolla" de dos hablas del área lingüística hispánica." *Thesaurus* 23: 193–205.

Laman, K. E. 1964 (1936). *Dictionnaire kikongo français*, 2 volumes. Ridgewood, NJ: The Gregg Press.

Lipski, J. 2012. "The 'new' Palenquero: Revitalization and Re-creolization." In *Varieties of Colombian Spanish*, ed. by R. File-Muriel, 21–41. Frankfurt and Madrid: Vervuert. https://doi.org/10.31819/9783954870196-003

Martínez, B. et al. 2017. "Ancestry estimates in afrodescendant population from San Basilio de Palenque, Colombia." *Forensic Science International. Genetics Supplement Series* 6 (2017): e224–e225. https://doi.org/10.1016/j.fsigss.2017.09.105

Martínez Miranda, G. et al. 2014. *Reglamento interno para la administración territorial del consejo comunitario "Ma Kankamaná" de San Basilio*. No place of publication: Consejo Comunitario Ma Kankamaná de San Basilio de Palenque.

Moñino, Y. 2003. "Lengua e identidad afro-americana: el caso del criollo de Palenque de San Basilio (Colombia)." In *Caminos cruzados. Ensayos en antropología social, etnoecología y etnoeducación*, ed. by C. Alès, and J. Chiappino, 517–531. Paris/Mérida (Venezuela): IRD-Editions/ULA-GRIAL. https://doi.org/10.4000/books.irdeditions.19094

Moñino, Y. 2017. "Past, Present, and Future of Palenquero Creole." In Schwegler, Kirschen, and Maglia (eds.), 15–50.

Moñino, Y. Forthcoming. "Un rare cas d'héritage syntaxique kikoongo en créole de San Basilio de Palenque (Colombie): les constructions à double objet nominaux et pronominaux." In *Du terrain à la théorie: Les 40 ans du Lacito Colloque international, 15–17 novembre 2016, Villejuif (France)*, ed. by I. Leblic, and L. Souag. Villejuif: Lacito Publications.

Morton, T. 2005. "Sociolinguistic Variation and Language Change in El Palenque de San Basilio (Colombia)". Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

Noguera, M. C.; Gusmão, L. et al. 2014. "Colombia's Population Crucible: Y Chromosome Evidence from Six Admixed Communities in the Department of Bolivar." *Annals of Human Biology* 41 (5): 453–459. E-print version available (Nov. 2013) at http://informahealthcare.com/eprint/tGM9Tr3eb239bP5R7Vzu/full https://doi.org/10.3109/03014460.2013.852244

Pérez Miranda, B. 2011. *Chitieno lengua ku ma kuendo*. Cartagena: Ediciones Pluma de Mompox S. A.

Pfleiderer, B. 1998 *Sprachtod und Revitalisierung der spanisch basierten Kreolsprache Palenquero (Kolumbien)* (Romanistische Linguistik, 7). Berlin: Freie Universität Berlin.

Schwegler, A. 1988. "Word-order Changes in Predicate Negation Strategies Revisited." *Diachronica* 5: 21–58. https://doi.org/10.1075/dia.5.1-2.03sch

Schwegler, A. 1991. "Negation in Palenquero: Synchrony." *Journal of Pidgin and Creole Languages* 6: 165–214. https://doi.org/10.1075/jpcl.6.2.02sch

Schwegler, A. 1994. "El Palenque de San Basilio (Colombia): Persistencia africana y problemas de (auto-)identificación de elementos lingüísticos subsaháricos." *Papia* 3: 6–30.

Schwegler, A. 1996. *"Chi ma nkongo": lengua y rito ancestrales en El Palenque de San Basilio (Colombia)*, 2 vols. Vervuert, Frankfurt / Iberoamericana, Madrid.

Schwegler, A. 2000. "The African Vocabulary of Palenque (Colombia). Part 1: Introduction and Corpus of Previously Undocumented Afro-Palenquerisms." *Journal of Pidgin and Creole Language* 15: 241–312. https://doi.org/10.1075/jpcl.15.2.02sch

Schwegler, A. 2002a. "El vocabulario africano de Palenque (Colombia). Segunda Parte: compendio de palabras (con etimologías)." In *Palenque, Cartagena y Afro-Caribe: historia y lingüística*, ed. by Y. Moñino, and A. Schwegler, 171–227. Tübingen: Max Niemeyer Verlag. https://doi.org/10.1515/9783110960228

Schwegler, A. 2002b. "On the (African) Origins of Palenquero Subject Pronouns." *Diachronica* 19 (2): 273–332. https://doi.org/10.1075/dia.19.2.03sch

Schwegler, A. 2006. "Bantu Elements in Palenque (Colombia): Anthropological, Archeological and Linguistic Evidence." In *African Re-genesis: Confronting Social Issues in the Diaspora*, ed. by J. B. Haviser, and K. C. MacDonald, 204–222. University College London Press, London.

Schwegler, A. 2007. "Black Ritual Insulting in the Americas: On the Art of 'Vociferar' (Colombia), 'Vacilar' (Ecuador) and 'Snapping', 'Sounding' or 'Playing the Dozens' (U.S.A.)." *Indiana* 24: 105–155. Available online at: http://www.iai.spk-berlin.de/publikationen/indiana/bisherige-ausgaben/indiana-24.html

Schwegler, A. 2011. "On the Extraordinary Revival of a Creole: Palenquero (Colombia)." In *Endangered Languages – Voices and Images*, ed. by M. Haboud, and N. Ostler, 153–165. Proceedings of FEL XV, Quito Ecuador, 7–9 September 2011. Bath, England: Foundation for Endangered Languages.

Schwegler, A. 2014. "Portuguese Remnants in the Afro-Hispanic Diaspora." In *Portuguese-Spanish Interfaces: Diachrony, Synchrony, and Contact*, ed. by A. M. Carvalho, and P. Amaral, 403–441. Amsterdam: John Benjamins.

Schwegler, A. 2016a. "Truth Reset: Pragmatics in Palenquero Negation." In *The Iberian Challenge: Creole Languages Beyond the Plantation Setting*, ed. by A. Schwegler, J. McWhorter, and L. Ströbel, 231–267. Vervuert, Frankfurt / Iberoamericana, Madrid. https://doi.org/10.31819/9783954878949-012

Schwegler, A. 2016b. "Combining Population Genetics (DNA) with Historical Linguistics: On the African Origins of Latin America's Black and Mulatto Populations." In *Spanish Language and Sociolinguistic Analysis*, ed. by S. Sessarego, and F. Tejedo, 33–88. Amsterdam: John Benjamins. https://doi.org/10.1075/ihll.8.02sch

Schwegler, A. 2017a. "On the African Origin(s) of Palenquero." In A. Schwegler, B. Kirschen, and G. Maglia (eds.), 51–119.

Schwegler, A. 2017b. "TED-x Talk: "African Roots in Latin America: Palenque (Colombia)." https://www.youtube.com/watch?v=HY4RTVaD9Sg&feature=youtu.be

Schwegler, A. 2018a. "Negation in Palenquero: Syntax, Pragmatics, and Change in Progress." In *Negation and Negative Concord in Creole Languages*, ed. by V. Deprez, and F. Henri, 257–288. Amsterdam: John Benjamins. https://doi.org/10.1075/coll.55.12sch

Schwegler, A. 2018b. "On the Controversial Origins of Non-Canonical Spanish and Portuguese Negation: Case Closed?" In *Lingua* 202: 24–43 (*Special Issue: Current trends in Afro-Hispanic linguistics*, ed. by S. Sessarego). https://doi.org/10.1016/j.lingua.2017.12.004

Schwegler, A., and A. Correa 2018. "Languages in Contact: The Case of Colombia." In *Biculturalism and Spanish in Contact: Sociolinguistic Case Studies*, ed. by E. Núñez, 145–175. New York: Routledge. https://doi.org/10.4324/9781315100357-8

Schwegler, A., Kirschen, B., and G. Maglia (eds.). 2017. *Palenque (Colombia): Orality, Identity, and Resistance. An Interdisciplinary Approach*. Amsterdam: John Benjamins. https://doi.org/10.1075/coll.54

Schwegler, A., and T. Morton 2003. "Vernacular Spanish in a Microcosm: *Kateyano* in El Palenque de San Basilio (Colombia)." *Revista Internacional de Lingüística Iberoamericana (RILI)* 1: 97–159.

Simarra, L., and Á. E. Triviño Doval 2012 (2007). *Gramátika ri luénga Palénque: Pa ma lo ke tan komensá [sic]*. Cartagena: Ediciones Pluma de Mompox S. A.

Simarra Obseso, R. 2006. "La lengua palenquera: una experiencia cosmovisionaria, significativa y creativa de los palenqueros descendientes de la diáspora africana en Colombia." *Anaconda (Bogotá)*, 8 (May 2008): 80–94. Available online at: https://issuu.com/educacionintercultural/docs/la_lengua_palenquera_como_simbolo_d

Simarra Obseso, R., R. Miranda Reyes, and J. Pérez Tejedor 2008. *Lengua ri Palenge. Jende suto ta chitiá*. Cartagena: Casa Editorial S. A.

Inflected infinitives in Galician

Michelle Sheehan[1], Jeffrey Blokzijl[2] & M. Carmen Parafita Couto[2]
[1]Anglia Ruskin University / [2]Leiden University

This study uses an audio-questionnaire, supplemented by spoken corpus data, to probe the acceptability of Galician inflected infinitives in different syntactic contexts. Our results reveal that inflected infinitives in Galician are acceptable in a different range of contexts than in closely related Portuguese, and also in a broader range of contexts than previously reported for this language. Intra-speaker variation appears to correlate most strongly with L1 (Spanish/Galician) and gender, with a trend for men and L1 Spanish speakers to be more accepting of the inflected infinitive.

Keywords: Galician, inflected infinitives, control, finiteness, gender

1. Introduction

Galician, like some varieties of Portuguese, Mirandese, Sardinian, Old Leonese and Old Neopolitan, has both inflected and uninflected infinitives (Gondar 1978; Longa 1994; Ledgeway 1998; Scida 2004). Inflected infinitives are morphologically marked in all person/number combinations except 1SG/3SG. Consider the paradigm for the irregular verb *ser* 'be': *ser, seres, ser, sermos, serdes, seren*. This verb form is mainly limited to subordinate clauses and differs from the subjunctive in being banned from finite clauses. Although it is widely known that Galician, like (European and Brazilian) Portuguese and unlike Spanish, has inflected infinitives, much less is known about the present usage of this verb form (but see Gondar 1978; Longa 1994; and Jansegers & Vanderschueren 2010). The present chapter aims to address this gap by presenting and discussing data from a spoken corpus and an audio survey (administered in 2017). Given the brevity of this chapter, we limit ourselves to a presentation of the survey and initial results, followed by a brief discussion of their potential significance. For a more comprehensive discussion and statistical analysis, see Sheehan, Shäfer & Parafita Couto (2019), which was actually written after this chapter.

Section 2 provides the background for the present study, including: an overview of the distribution of inflected infinitives in Brazilian and European Portuguese (Section 2.1); a survey of normative and prescriptive pressures on the use of Galician

https://doi.org/10.1075/rllt.16.16she

inflected infinitives (Section 2.2); and a discussion of attestations of inflected infinitives in the *Corpus Oral Informatizado da Lingua Galega* (CORILGA) (Section 2.3). Section 3 lays out the methodology of the present study and Section 4 discusses our results and their implications. Finally, Section 5 concludes the paper.

2. Background to the present study

2.1 The distribution of inflected infinitives in Portuguese

Given their shared history, one might expect the distribution of inflected infinitives in Galician to mirror that of the equivalent grammatical phenomenon in European/Brazilian Portuguese. For this reason, we begin this chapter with a brief overview of the distribution of Portuguese inflected infinitives, highlighting controversial issues and apparent points of variation and change.

Most work on inflected infinitives in European Portuguese (EP) has highlighted the contexts where they occur with a full referential subject, for example in extraposed subject clauses (see Raposo 1987):

(1) *Será* *difícil* *[(eles)* *aprovarem* *a* *proposta].*
be.FUT.3SG difficult they approve.INF.3PL the proposal
'It will be difficult for them to approve the proposal.' [EP, Raposo (1987: 86)]

In EP, at least the following contexts permit overt/covert referential subjects of this kind: complements of epistemic/factive verbs, complements of nouns, and adjunct clauses (see Raposo 1987). Brazilian Portuguese (BP) also permits inflected infinitives with referential subjects, though this is subject to substantial inter-speaker variation and may have a different status from the superficially similar phenomenon in EP (see Modesto & Maia 2017).

Inflected infinitives are also claimed to surface in instances of non-obligatory control in both EP and BP (Pires 2001), whereby the subject of an inflected infinitive is bound by a linguistic antecedent that is not necessarily local. In the following example, for instance, either linguistic antecedent can bind the embedded subject, showing that this cannot be an instance of obligatory control:

(2) *Os professores*$_i$ *disseram aos* *alunos*$_j$ *que*
the teachers said.3PL to.the students that
era *preciso* [pro$_{i/j}$ *trabalharem]*
was necessary work.INF.3PL.
'The teachers told the students that it was necessary to work.'

The status of inflected infinitives in obligatory control contexts is more controversial, however (see Landau 2015). Inflected infinitives are undeniably possible

under 'control' verbs such as *prometer* 'promise' and *persuadir* 'persuade' in EP and BP (see Raposo 1989; Madeira 1994):

(3) a. EP exhaustive object OC (Sheehan 2014b: 120)
O professor persuadiu os alunos a fazer(em) o trabalho.
the teacher persuaded the pupils A do.INF.3PL the work
'The teacher persuaded the pupils to do the work.'
[uninflected 70%; inflected 70%, *n* = 44]

b. EP partial object OC (Sheehan 2014b: 120)
Os professores persuadiram o diretor
the teachers persuaded the headteacher
a reunir(em)=se mais tarde.
A meet.INF.3PL=se + late
'The teachers persuaded the headteacher to meet later on.'
[uninflected 58%; inflected 95%, *n* = 19]

Moreover, in such contexts, these null subjects have been argued to have at least some of the properties of control (see Modesto 2010; Modesto & Maia 2017; Sheehan 2014b, 2018).

It has recently been proposed, however, that apparent instances of obligatorily controlled inflected infinitives are illusory (Landau 2016), or actually involve coercion of control predicates into non-control co-reference (Barbosa 2018). A relevant aspect of this debate is the fact that inflection is more widely accepted in instances of partial as opposed to exhaustive control in both varieties of Portuguese. Exhaustive subject control refers to obligatory co-reference between the subject of an infinitive and the subject of the next clause up, whereas partial control denotes contexts where the controller represents a subset of the infinitival subject. Inflection is generally worse in instances of exhaustive control and better in instances of partial control (Sheehan 2014b, 2018), but, for many speakers, inflection is not obligatory even in partial control contexts (but see Sheehan 2014a for discussion of this):

(4) a. EP exhaustive non-local subject OC (Sheehan 2014b: 119)
Prometemos à professora chegar(%mos) a tempo.
promised to.the teacher arrive.INF.1PL at time
'We promised the teacher to arrive on time.'
[uninflected 95%; inflected 37%, *n* = 19]

b. EP partial non-local subject OC (Sheehan 2014b: 119)
O Pedro prometeu à Ana reunir(em)=se em Braga.
the Pedro promised to.the Ana meet.3PL=SE in Braga
'Pedro promised Ana to meet in Braga.'
[uninflected 76%; inflected 90%, *n* = 19]

In cases of local subject control, with verbs like *preferir* 'prefer', no speakers accept inflection in cases of exhaustive control in either variety, though examples of this kind are found in written corpora (Gonçalves, Santos & Duarte 2014).

(5) EP exhaustive local subject OC (Sheehan 2014b: 118)
Preferíamos receber(*mos) *um salário maior.*
prefer.1PL receive.INF.1PL a salary higher
'We would prefer to get a higher salary.'
[uninflected 100%; inflected 0%, *n* = 19]

Whatever the formal status of these subjects, it seems that inflected infinitives are acceptable for many EP and BP speakers in complement domains with either referentially free or controlled subjects, except in instances of exhaustive local subject control, where inflection is simply banned (possibly due to obviation – Sheehan 2014b).

Although there is substantial variation across EP and BP speakers (see Modesto & Maia 2017; Sheehan 2018 for discussion), in general, it has been reported that inflected infinitives are used more in Portuguese than in Galician. Previous studies report that inflected infinitives can occur only in subject and adverbial clauses, apositions, predicative complements and the complements of nouns and adjectives, with their appearance in the complements of verbs severely restricted (Gondar 1978; Jansegers & Vanderschueren 2010). In fact, in actual usage, inflected infinitives have been reported to be mainly restricted to adverbial clauses, due to Spanish influence (Jansegers & Vanderschueren 2010: 2, citing Gondar 1978; Freixeiro Mato 2002: 389-396). In the following section, we review prescriptive pressures on Galician and argue that these largely reflect reported usage, before considering data from the spoken corpus and judgements, which suggest a wider range of contexts of attestation.

2.2 Normative pressures and prescriptive norms

After becoming an official language in Galicia under the 1978 Spanish constitution, Galician moved quickly through the process of standardisation (Santamarina Fernández 1995; Ramallo & Rei-Doval 2015), with a written standard being proposed in 1983, made law and then implemented (Kabatek 1997). Nonetheless, prescriptive norms of many aspects of grammar remain unclear, partly because the *Real Academia Galega*, established in 1906, has not yet published an official Galician grammar (Álvarez Blanco, Cidrás Escáneo, González Seoane, Regueira Fernández & Xove Ferreiro 2004). It is therefore difficult to know what the prescriptive norms are regarding the use of inflected infinitives in Galician.

Classroom materials focus explicitly on the existence of this morphological form but are rather vague in terms of its distribution and function. One popular resource claims that the infinitive "will only ever be inflected whenever needed for reasons of clarity, as inflection is uneconomical". This is defined as being "[w]henever the infinitive, as well as its own subject, can also have other subjects corresponding to other grammatical persons".[1] The implication here is that inflection serves a purely functional purpose, being used to indicate the person/number features of the verb, in contexts of potential ambiguity. As we shall see below, this is seemingly not the function it has for many of the respondents to our questionnaire, who permit inflection in instances of what looks like obligatory control (where there is co-reference between the subject of an inflected infinitive and an argument in the next clause up (6)). Moreover, inflection is also permitted wherever an infinitive has an overt subject, in which case the person and number features of the subject are wholly unambiguous (as in (7), based on the native judgment of the third author):

(6) *Os mozos prometéronlle ás mozas non beberen de máis.*
The guys promised=3DAT the.DAT girls not drink.INF.3PL of much
'The guys$_1$ promised the girls PRO$_1$ not to drink too much.'

(7) *Será difícil [aprobar(en) eles a proposta].*
be.FUT.3SG difficult approve.(INF.3PL) they the proposal
'It will be difficult they to-approve-Agr the proposal.'

This normative description is thus, at best, insufficient to describe the intuitions of native speakers, at worse, highly misleading.

The kinds of exercises used in conjunction with these teaching materials involve either the simple conjugation of infinitives in example sentences where the context is given as an appropriate one or, more difficult examples where the student must also decide whether inflection is appropriate in a given syntactic context. An analysis of two of the answer keys to such sets of exercises, comprising a total of 74 infinitives of which 65 can be inflected is highly

1. *¿Cando logo se poderá conxugar o infinitivo? Conxugarase sempre que por expresividade se considere oportuno, xa que non é un acto de fala económico. ¿Cando é oportuno expresivamente? Sempre que o infinitivo ademais de poder presentar o suxeito que ten, poida ter outros suxeitos, outros suxeitos correspondentes a outras persoas gramaticais* https://gl.wikibooks.org/wiki/Curso_de_lingua_galega/O_verbo/O_infinitivo_conxugado

revealing.[2] Of the 65 grammatical contexts, 48 (74%) occur in adjunct clauses and 9 (14%) in extraposed subject clauses. The remainder of the grammatical contexts are complements of nouns (6%), two obligatory control examples with peculiar properties, a matrix command and a complement clause. These example sentences are largely in line with what is reported by Gondar (1978) and others, as discussed above, suggesting that inflected infinitives are mainly restricted to adjunct clauses, but also possible in the complements of nouns and in extraposed subject clauses.

A consideration of the 11 ungrammatical contexts is arguably also revealing, however. Most of them (64%) involve exhaustive local subject control contexts of the following kind, where the use of inflection is apparently proscribed (the 'correct' answer without inflection is given here):

(8) *Debiades* ***explicalo*** *mellor.*
Must. IMP.2PL explain.INF.it better
'You should explain it better.'

It is particularly interesting that the proscribed contexts mainly involve obligatory control, because as we shall see below, many of our respondents actually accept inflection in this context, even in what look like exhaustive control contexts. It is possible that the proscription of inflection in this context is essentially a response to recent sociolinguistic change, which has seen the expansion of the inflected infinitive into obligatory control contexts.

Despite the fact that the grammatical properties of the inflected infinitives are poorly described in teaching materials, speakers are certainly aware of its existence as a feature of Galician grammar. Nonetheless, as Kabatek (1997) notes, Gondar (1978: 55), in his description of spoken Galician in the 1970s, claimed that inflected infinitives were a moribund feature of spoken Galician grammar, being much reduced in their distribution, possibly under the influence of Castilian Spanish, which lacks such a form (see Quicoli 1996). Interestingly, though, Gondar also notes an increase in the use of inflected infinitives in the formal writing of some of his contemporaries attributing it to a desire to *recuperar* 'get back' the inflected infinitive and *evitar a súa perda* 'avoid its loss' (Gondar 1978: 139–140). An interesting question is whether the use of inflected infinitives might also have spread in spoken Galician due to similar

2. https://gl.wikibooks.org/wiki/Exercicios_de_morfosintaxe_en_lingua_galega/verbo/Infinitivo_conxugado http://www.ogalego.eu/exercicios_de_lingua/exercicios/morfoloxia/verbo2.htm#9

sociolinguistic pressures, notably the desire of Galician speakers, conscious or otherwise, to distinguish their language from Spanish. The aim of our ongoing study is to investigate this possibility.

2.3 Attestation in spoken corpora

Or course, normative pressures often do not correlate with usage or native intuitions. For this reason, in constructing our questionnaire, we also drew on examples of inflected infinitives in the *The Corpus Oral Informatizado da Lingua Galega* (CORILGA). This spoken corpus consists of 98 hours of informal and formal conversations recorded from 1960 onwards (see García-Mateo, Cardenal, Regueira, Fernández Rei, Martinez, Seara, Varela & Basanta 2014). There are only 71 examples of inflected infinitives in the corpus, in line with what has been claimed about the low frequency of the form. While they are found most frequently in: (1) adjuncts (n = 44, 62%); and (2) complements of nouns (n = 10, 14%), as we might expect given the discussion above, there are also more rare examples in complements of Adj (n = 3, 4%); extraposed subject clauses (n = 2, 3%) and even in exhaustive local subject control contexts (n = 3, 4%). The final context is of particular interest as this is a context in which inflection is proscribed in Galician (see above) and banned in both European and Brazilian Portuguese (though examples can be found in written corpora). There are three such examples in the corpus, all attested after 2002, for example:

(9) *Esperamos* | *tamén* | *non* ***sermos*** *os* *únicos* *nesta* *cámara* *como*
hope.1P also NEG be.1PL the only in.this room like
noutras *ocasións.*
on.other occasions
'We also hope not to be the only ones in this room like on other occasions'
(OFDL-CORILGA-AYMERICH-02-2009)

Given the attestation of such examples, albeit in very small numbers, in conjunction with the classroom materials discussed above, it is interesting to know what speakers' intuitions are regarding inflection in these exhaustive 'control' contexts, which disallow inflection in Portuguese. Given that all three attested examples of a controlled inflected infinitive in the corpus and all the proscribed examples in teaching materials involve exhaustive local subject control, it is also interesting to establish the patterns for object control and non-local subject control, as well as for partial control, given the variation attested in varieties of Portuguese, as discussed above.

3. Methodology

With the aim of describing the intuitions of native speakers of Galician regarding the acceptability of inflected infinitives across a range of syntactic contexts, we designed an online audio questionnaire consisting of 50 target sentences, 24 fillers, and 15 social profiling questions. The target sentences consisted of an inflected infinitive in one of the following syntactic contexts, with at least three structurally equivalent examples per condition.[3] The syntactic contexts were: adjunct clause, complement of noun, extraposed subject clause, local subject control (exhaustive/partial), non-local subject control (exhaustive/partial), object control (exhaustive/partial), factive (partial control), factive (non-control), epistemic (partial control), epistemic (non-control), desiderative (non-control).[4] The contexts were chosen based on the corpus examples, the descriptive and prescriptive literature on Galician and Portuguese, and the native intuitions of the third author. We attempted to include most contexts where inflection might be possible. All sentences were recorded by a native speaker of Galician (not an author of this paper), and embedded in a Qualtrics questionnaire, mixed with fillers. They appeared in randomized order and were rated on a 5-point scale, using emoticons.[5] A total of 314 participants filled in the questionnaire and participation was unpaid. Initially, respondents were found through personal and academic networks, but after an article was published in a local Galician newspaper, larger numbers of people filled in the questionnaire who are unknown to the researchers.

Of the 314 participants, 189 were female and 125 male. The average age was 37.04, ranging from 16 to 81 years of age. All participants were self-identified as early sequential bilinguals: 236 Galician mother tongue (i.e. they learned Galician first, and Spanish later, but both languages were learned during childhood), 78 Spanish mother tongue (they learned Spanish first, followed by Galician, also during childhood).[6]

---

3. We constructed the examples in order to make them as natural as possible, basing ourselves on the attested examples in Gondar (1978) or the CORILGA corpus, and avoiding overt pronominal subjects in order to avoid issues of subject placement.

4. Because examples with auxiliary verbs often lead to greater acceptability in European Portuguese, we also added examples with auxiliary verbs in later permutations of the survey, but we saw no effect of the presence or absence of an auxiliary in the complements of epistemic or factive predicates, so we collapse this distinction here.

5. https://www.qualtrics.com

6. Most participants in our survey came from the province of A Coruña (159/314). Given the uneven geographical distribution, we are not able to test whether geography is a relevant sociolinguistic factor.

4. Results and discussion

The crude results show that there are contexts in which inflected infinitives are acceptable for almost all speakers, contexts where they are not and more variable contexts, notably most control contexts, but also extraposed subject clauses. For the full data set see and Parafita Couto, Sheehan, Blokzijl & Shäfer 2019. We include just one control context here, by way of illustration.

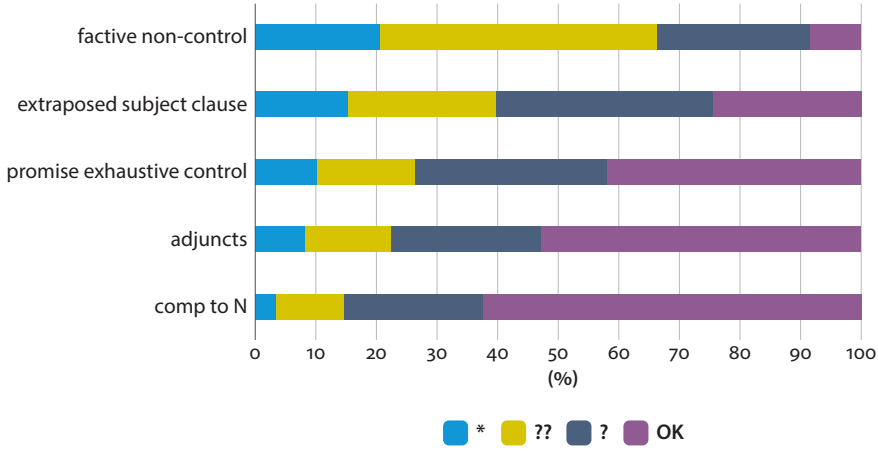


Figure 1. Basic results across contexts

Unsurprisingly, the vast majority of speakers strongly or more marginally accepted inflection in the complements of nouns (85%) and clausal adjuncts (77%), mirroring both the most frequent corpus examples and also the prescriptive norms:

(10) *Non se che pode dar o dereito de escolleres sempre.*
NEG know what can give.INF the right of choose.INF.2SG always
'I don't know what always gives you the right to choose.'
[imean = 3.9; cmean = 3.8][7]

(11) *Para quedares na casa, tiñas que estar moi* for *enfermo.*
remain.INF.2SG in.the house had.2SG that be.INF very ill
'For you to stay at home, you must be really ill.'
[imean = 4.0; cmean = 3.7]

7. We give the item mean (imean) and context mean (cmean) for each sentence based on a numerical 5-point scale. When translating individual cmeans into grammaticality judgments, we take a cmean of 0–1.9 to be *; 2–2.9 to be ??; 3–3.9 to be ? and 4–5 to be OK. Note that because of the substantial variability in judgements, the imean for a particular example can look lower than expected and a middling cmean can be due to either general low acceptability or to high variability. The data in Table 1 is therefore a better indication of attested variation than the numbers alone.

(12) *Para sanares tes que tomar este xarabe.*
for heal.INF.2SG have.2SG to drink this medicine.
'For you to heal, you have to drink this medicine.'
[imean = 3.6; cmean = 3.7]

Most speakers uniformly rejected uncontrolled subjects (with epistemic, factive and desiderative verbs with or without an auxiliary verb) (see Longa 1994, and cf. European Portuguese, Raposo 1987). For factive verbs, 67% of respondents rejected examples like the following, for example:

(13) **Lamento teren perdido os documentos.*
regret.1SG have.INF.3PL lost the.PL documents
[imean = 2.5; cmean = 2.5]

These kinds of examples can be contrasted with examples involving partial co-reference between the matrix subject and embedded subject, which we label partial control, for ease of reference:

(14) %*Lamento termos perdido os documentos.*
regret.1SG have.INF.1PL lost the documents
'I regret that we have lost the documents.' [imean = 3.4; cmean = 3.4]

Such examples are much more acceptable than non-control examples, suggesting that in complement domains, the subjects of inflected infinitives cannot be referentially free, unlike in (Brazilian and European) Portuguese.[8] With epistemic and desiderative verbs, there is only a slight improvement if the subject of the matrix clause is a potential partial controller of the embedded subject (2.8 vs. 1.8 cmean). It seems that epistemic predicates cannot take an inflected infinitival complement for most speakers, though where they can, control is required.

For many speakers, inflected infinitives are also grammatical in instances of object control (*avisar* 'notify', *convencer* 'convince', *forzar* 'force') and non-local subject control (*prometer* 'promise', *xurar* 'swear', *ofrecer* 'offer'). Note that there is apparently no sensitivity to the exhaustive/partial control distinction here (in contrast with what was observed above for Brazilian and European Portuguese).

8. We also have not yet tested the possibility of partial control readings with uninflected infinitives in Galician (see Sheehan 2014a on this issue in European Portuguese). Note that due to nature of this survey, we also haven't tested whether such examples have the diagnostic properties of obligatory control. The contrast between non-control and partial control is however suggestive in this respect.

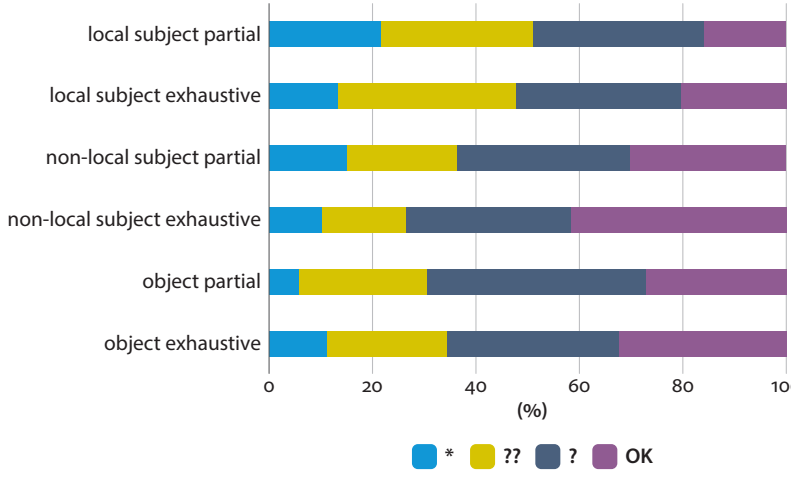


Figure 2. Irrelevance of partial vs. exhaustive control

While inflection is more acceptable in instances of non-local subject control and object control, in specific contexts, there is no major difference in the acceptability of inflection in exhaustive vs. partial control contexts. The one slight exception is non-local subject control, where inflection is slightly more acceptable in *exhaustive* control contexts, in direct contrast with what we saw above for European Portuguese. This is an interesting finding as it clearly shows that Galician differs from Portuguese with respect to this pattern.

In fact, while they are less acceptable, inflected infinitives were surprisingly good for some speakers in exhaustive local subject control contexts (unlike in Portuguese). Recall that there were also some corpora examples of this kind. Such examples are by no means fully grammatical for all speakers, but they were accepted or marginally accepted by 52% of the respondents:

(15) %*Preferimos quedarmos aquí hoxe.*
prefer.1PL stay.INF.1PL here today
'We prefer to stay here today.' [imean = 3.0; cmean = 2.8]

One potential analysis of these patterns is that inflection on Galician infinitives is freely available for some speakers. This is not generally the case, however, for many Galician speakers, as our survey shows. Rather, the results show that most speakers have intuitions regarding which contexts permit inflection and which do not, but that some speakers are more permissive than others. Nonetheless, we need more data regarding the distribution of uninflected infinitives to understand the status of restrictions on inflected infinitives with referential subjects. Note that functional explanations in terms of disambiguation are problematic as inflection

is widely accepted in exhaustive control contexts, where there is only one possible referent for the subject. So what determines how permissive Galician speakers are in this respect?

Our initial analysis suggests that variation across syntactic contexts correlates with two factors: gender and mother tongue.[9] Figure 3 below, shows variation according to gender. The number of male participants is substantially lower than the number of female participants, but the initial results of the survey suggest that men appear to be more accepting of inflected infinitives than women are in a range of contexts, including in syntactic contexts (e.g. desiderative non-control) where most participants did not accept inflection.

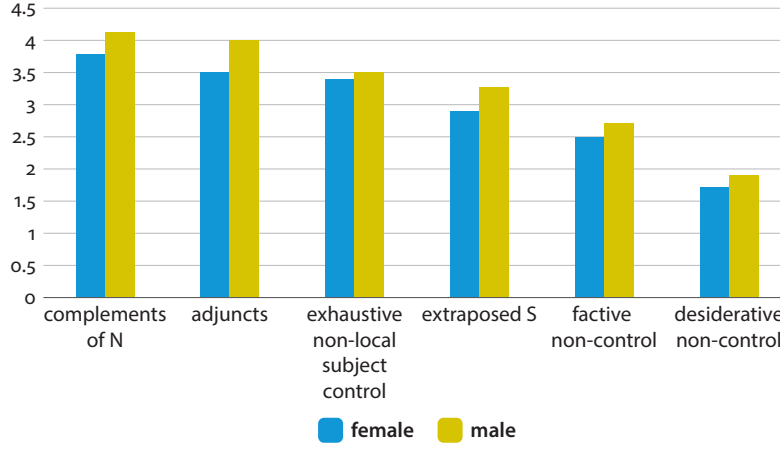


Figure 3. Variation according to gender

Interestingly, not only are men less likely to reject inflected infinitives in almost all contexts, including the least widely accepted ones, they are also overrepresented in the CORILGA corpus examples, where 72% of inflected infinitives are uttered by men.[10]

A similar trend is observed when we consider the mother tongue (language learned first) of these sequential bilinguals. It appears that self-declared Spanish

9. Of the Spanish mother tongue speakers, we have 34 participants who are male and 44 who are female.

10. The CORILGA corpus includes data from 460 men and 290 women, so 61% of speakers are male, though we do not know what percentage of utterances come from males. At the time this study was conducted, transcription was not complete, so this percentage is based on the available transcripts.

mother tongue speakers rate the acceptability of inflected infinitives slightly higher than Galician mother tongue speakers in almost all contexts, as can be seen in Figure 4.

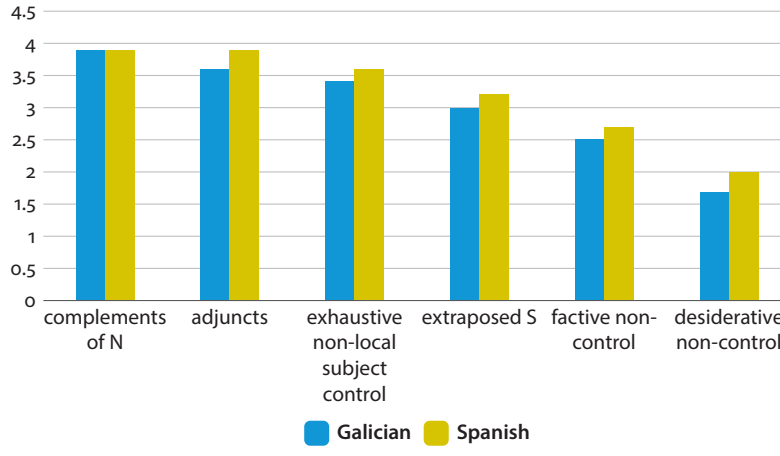


Figure 4. Variation according to mother tongue

We propose that these patterns may be due to prescriptive norms that are enforcing the use of the inflected infinitives as a marker of the Galician language. Nonetheless, we need to test the statistical validity of these claims, ideally with more balanced numbers of male/female and Spanish/Galician mother tongue speakers or else with more sophisticated statistical analysis. We take up these issues in ongoing work.

5. Conclusion

The purpose of this study was to document, for the first time, the acceptability of inflected infinitives in Galician based on a large-scale audio survey. Although the prescriptive norms regarding the use of inflected infinitives are not completely clear, classroom materials focus on the existence of this morphological form and try to explain its use in terms of ambiguity resolution. Our data suggest that these classroom characterisations are inadequate and that the syntactic contexts in which inflected infinitives are now accepted are much broader than in the normative literature.

There is, however, a great deal of inter-speaker variation regarding acceptability. While in islands (e.g. adjunct clauses, complements of N) inflected infinitives are clearly grammatical for most Galician bilinguals, in control contexts, they are

less consistently accepted. This variation seems to correlate most strongly with mother tongue and gender: we observed a trend for Spanish mother tongue speakers to be generally more accepting of these contexts than Galician mother tongue speakers, and for men to be more accepting of the inflected infinitive than women. Taken into account that Spanish lacks inflected infinitives and following Kabatek (1997), we propose that this verbal form acts as a strong marker of regional identity. This factor may have led to these two groups extending the use of the inflected infinitive to syntactic contexts outside the traditional ones.

References

Álvarez Blanco, Rosario, Franciso A. Cidrás Escáneo, Ernesto X. González Seoane, Xosé L. Regueira Fernández, and Xosé Xove Ferreiro. 2004. "Entre o uso e a norma. O proxecto da Gramática da Real Academia Galega." In *Norma lingüística e variación: Unha perspectiva desde o idioma galego*, ed. by Rosario Álvarez, and Henrique Monteagudo, 61–68. Santiago de Compostela: Consello da Cultura Galega/Instituto da Lingua Galega.

Barbosa, Pilar. 2018. "'Partial Control' Inflected Infinitives are not Obligatorily Controlled." Paper presented at *Olomouc Linguistics Colloquium* (Olinco). Olomouc, Czechia: Palacký University.

Freixeiro Mato, Xosé R. 2002. *Lingua galega: normalidade e conflito*. Santiago de Compostela: Laiovento.

García-Mateo, Carmen, Antonio Cardenal, Xosé L. Regueira, Elisa Fernández Rei, Marta Martinez, Roberto Seara, Rocío Varela, and Noemí Basanta. 2014. "CORILGA: A Galician Multilevel Annotated Speech Corpus for Linguistic Analysis." *9th Language Resources and Evaluation Conference* (LREC 2014), 26–31. Reykjavik.

Gonçalves, Anabela, Ana-Lúcia Santos, and Inês Duarte. 2014. "(Pseudo)inflected infinitives and control as Agree." In *Romance Languages and Linguistic Theory: Selected papers from 'Going Romance' Leuven*, ed. by Lahousse, Karen, and Stephanie Marzo, 161–180. Amsterdam: John Benjamins.

Gondar, Francisco G. 1978. *O infinitivo conxugado en galego*. Santiago de Compostela: Universidade de Santiago.

Jansegers, Marlies, and Clara Vanderschueren. 2010. "El infinitivo flexionado gallego: ¿entre portugués y castellano? *Revue de Linguistique Romane* 74: 415–441.

Kabatek, Johannes. 1997. "Strengthening Identity: Differentiation and Change in Contemporary Galician." In *Taming the Vernacular. from Dialect to Written Standard Language*, ed. by Cheshire, Jenny, and Dieter Stein, 185–199. London/New York: Longman.

Landau, Idan. 2015. *A Two-Tiered Theory of Control*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262028851.001.0001

Landau, Idan. 2016. "Agreement at PF: An Argument from Partial Control." *Syntax* 19 (1): 79–109. https://doi.org/10.1111/synt.12118

Ledgeway, Adam. 1998. "Variation in the Romance Infinitive: The Case of the Southern Calabrian Inflected Infinitive." *Transactions of the Philological Society Volume* 96 (1): 1–61. https://doi.org/10.1111/1467-968X.00023

Longa, Víctor M. 1994. "The Galician Inflected Infinitive and the Theory of UG." *Catalan Working Papers in Linguistics* 4 (1): 23–44.

Madeira, Ana. 1994. "On the Portuguese Inflected Infinitive." *UCL Working Papers in Linguistics* 6: 179–203.

Modesto, Marcello. 2010. "What Brazilian Portuguese says about Control: Remarks on Boeckx & Hornstein." *Syntax* 13 (1): 78–96. https://doi.org/10.1111/j.1467-9612.2009.00136.x

Modesto, Marcello, and Marcus Maia. 2017. "Representation and Processing of the Inflected Infinitive in Brazilian Portuguese: An Eye-Tracking Study." *Revista de Estudos da Linguagem* 25 (3): 1183–1224. https://doi.org/10.17851/2237-2083.25.3.1183-1224

Parafita Couto, Maria Carmen, Michelle Sheehan, Jeffrey Blokzijl, and Martin Schäfer. 2019. Galician inflected infinitives: A dataset containing grammaticality judgements and sociolinguistic factors. figshare. Dataset. https://doi.org/10.6084/m9.figshare.8026922.v1

Pires, Acrisio. 2001. "PRO, Movement and Binding in Portuguese." In *Romance Syntax, Semantics and L2 Acquisition: Selected Papers from the 30th Linguistic Symposium on Romance Languages*, ed. by Camps, Joaquim, and Caroline R. Wiltshire, 153–168. Amsterdam: John Benjamins.

Quicoli, Antonio C. 1996. "Inflection and Parametric Variation: Portuguese vs. Spanish." In *Current Issues in Comparative Grammar*, ed. by Freidin, Robert, 46–100. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-009-0135-3_2

Ramallo, Fernando, and Gabriel Rei-Doval. 2015. "The Standardization of Galician." *Sociolinguistica* 29: 61–82. https://doi.org/10.1515/soci-2015-0006

Raposo, Eduardo. 1987. "Case Theory and Infl-to-Comp: The Inflected Infinitive in European Portuguese." *Linguistic Inquiry* 18 (1): 85–109.

Raposo, Eduardo. 1989. "Prepositional Infinitival Constructions in European Portuguese." In *The Null Subject Parameter*, ed. by Osvaldo Jaeggli, and Kenneth Safir, 277–305. Dordrecht: Kluwer.Spanish. https://doi.org/10.1007/978-94-009-2540-3_10

Santamarina Fernández, Antón. 1995. "Norma e estándar." In *Estudios de sociolingüística galega sobre a norma do galego culto*, ed. by Monteagudo, Henrique, 53–98. Vigo: Galaxia.

Scida, Emily. 2004. *The Inflected Infinitive in Romance Languages*. New York/London: Routledge. https://doi.org/10.4324/9780203488768

Sheehan, Michelle. 2014a. "Partial Control in Romance Languages: The Covert Comitative Analysis." In *Romance Languages and Linguistic Theory*, ed. by Lahousse, Karen, and Stephanie Marzo, 181–198. Amsterdam: John Benjamins.

Sheehan, Michelle. 2014b. "Portuguese, Russian and the Theory of Control." In *Proceedings of the 43rd Annual Meeting of the North East Linguistic Society 43 (NELS 43)*, volume 2, ed. by Huang, Hsin-Lun, Ethan Poole, and Amanda Rysling, 115–126. Amherst, MA: GLSA.

Sheehan, Michelle. 2018. "Control of Inflected Infinitives in European Portuguese." In *Complement clauses in Portuguese: syntax and acquisition*, ed. by Gonçalves, Anabela, and Ana-Lúcia Santos, 29–58. Amsterdam: John Benjamins.

Sheehan, Michelle, Maria Carmen Parafita Couto, and Martin Schäfer. 2019. Crowdsourcing and minority languages: The case of Galician inflected infinitives. *Frontiers in Psychology – Language Sciences* 10: 1157.

Index

The chapters in this book represent the theme of "bridges" – bridging research approaches and directions across languages, methodologies and disciplines. Alongside descriptive and theoretical studies, the contributions present experimental studies addressing issues in syntax, phonetics-phonology and sociolinguistics. And alongside investigations of linguistic phenomena in standard Romance language varieties, other investigations address less well-known and studied, minority and endangered varieties (e.g., Quebec French, Brazilian Portuguese, Romanian, Galician, Catalan and Palenquero) from both synchronic and diachronic perspectives. Romance languages in contact with other languages and bilingualism, now also integral aspects of the field, are reflected in this volume as well, including less well-known cases of contemporary contact of Serbian with Romanian, and earlier contact of African languages with Spanish and Portuguese. This volume thus continues the decades long tradition of the Linguistic Symposium on Romance Languages of embracing cutting-edge developments in the field.

ISBN 978 90 272 0555 1
9 789027 205551

John Benjamins Publishing Company