

Studies in Historical Ibero-Romance Morpho-Syntax

Edited by
Miriam Bouzouita,
Ioanna Sitaridou
and Enrique Pato

John Benjamins Publishing Company

Studies in Historical Ibero-Romance Morpho-Syntax

Issues in Hispanic and Lusophone Linguistics (IHLL)

ISSN 2213-3887

IHLL aims to provide a single home for the highest quality monographs and edited volumes pertaining to Hispanic and Lusophone linguistics. In an effort to be as inclusive as possible, the series includes volumes that represent the many sub-fields and paradigms of linguistics that do high quality research targeting Iberian Romance languages. IHLL considers proposals that focus on formal syntax, semantics, morphology, phonetics/phonology, pragmatics from any established research paradigm, as well as psycholinguistics, language acquisition, historical linguistics, applied linguistics and sociolinguistics. The editorial board is comprised of experts in all of the aforementioned fields.

For an overview of all books published in this series, please see <http://benjamins.com/catalog/ihll>

Editors

Jennifer Cabrelli Amaro
University of Illinois at Chicago

Kimberly L. Geeslin
Indiana University

Editorial Board

Patrícia Amaral
Indiana University

Sonia Colina
University of Arizona

João Costa
Universidade Nova de Lisboa

Inês Duarte
Universidade de Lisboa

Daniel Erker
Boston University

Timothy L. Face
University of Minnesota

Sónia Frota
Universidade de Lisboa

Ángel J. Gallego
Universitat Autònoma de Barcelona

María del Pilar García Mayo
Universidad del País Vasco

Anna Gavarró
Universitat Autònoma de Barcelona

Michael Iverson
Indiana University

Matthew Kanwit
University of Pittsburgh

Paula Kempchinsky
University of Iowa

Naomi Lapidus Shin
University of New Mexico

Juana M. Licerias
University of Ottawa

John M. Lipski
Pennsylvania State University

Gillian Lord
University of Florida

Jairo Nunes
Universidade de São Paulo

Acrisio Pires
University of Michigan, Ann Arbor

Pilar Prieto
Universitat Pompeu Fabra

Liliana Sánchez
Rutgers University

Ana Lúcia Santos
Universidade de Lisboa

Scott A. Schwenter
Ohio State University

Carmen Silva-Corvalán
University of Southern
California

Miquel Simonet
University of Arizona

Megan Solon
State University of New York

Juan Uriagereka
University of Maryland

Elena Valenzuela
University of Ottawa

Bill VanPatten
Michigan State University

Volume 16

Studies in Historical Ibero-Romance Morpho-Syntax

Edited by Miriam Bouzouita, Ioanna Sitaridou and Enrique Pato

Studies in Historical Ibero-Romance Morpho-Syntax

Edited by

Miriam Bouzouita

Ghent University

Ioanna Sitaridou

University of Cambridge, Queens' College

Enrique Pato

Université de Montréal

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/ihll.16

**Cataloging-in-Publication Data available from Library of Congress:
LCCN 2017060210 (PRINT) / 2017061464 (E-BOOK)**

ISBN 978 90 272 0045 7 (HB)

ISBN 978 90 272 6431 2 (E-BOOK)

© 2018 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

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

Table of contents

CHAPTER 1	
Some introductory reflections	1
<i>Miriam Bouzouita, Ioanna Sitaridou & Enrique Pato</i>	
CHAPTER 2	
Velar allomorphy in Ibero-Romance: Roots, endings and clashes of morphemes	13
<i>Paul O'Neill</i>	
CHAPTER 3	
The history of concatenative compounds in Spanish	47
<i>María Irene Moyna</i>	
CHAPTER 4	
Intersubjectification and textual emphasis in the use of <i>definite article</i> + <i>proper name</i> in Spanish	75
<i>Miguel Calderón Campos</i>	
CHAPTER 5	
Stylistic fronting in Old Spanish texts	99
<i>Javier Elvira</i>	
CHAPTER 6	
The dative experiencer of Spanish <i>gustar</i>	123
<i>Chantal Melis & Marcela Flores</i>	
CHAPTER 7	
Postverbal subjects of unaccusative verbs in the history of Portuguese	149
<i>Kristine Gunn Eide</i>	
CHAPTER 8	
On the position of overt subjects in infinitival clauses in Spanish and Portuguese: Pragmatic, semantic and frequency-based motivations	173
<i>Kim Schulte</i>	

CHAPTER 9

- Allative to purposive grammaticalisation: A quantitative story of Spanish *para* 195
Rena Torres Cacoullas & Joseph Bauman

CHAPTER 10

- Recurrent processes in the evolution of concessive subordinators
in Spanish and Catalan 223
Manuel Pérez Saldanya & José Ignacio Hualde

CHAPTER 11

- Si* as a Q-particle in Old Spanish 249
Javier Rodríguez Molina & Andrés Enrique-Arias

CHAPTER 12

- Realmente, verdaderamente* and *ciertamente*: On the relation between
epistemic modality and intensifying adverbs 275
Mónica González Manzano

CHAPTER 13

- Variation and the use of discourse markers in 16th-century Spanish 303
Christopher J. Pountain

CHAPTER 14

- Studying Ibero-Romance before 1200 325
Roger Wright

- Index 337

Some introductory reflections

Miriam Bouzouita, Ioanna Sitaridou & Enrique Pato
Ghent University / University of Cambridge / Université de Montréal

1. General overview

As the title of this chapter indicates, we will offer some introductory reflections on the contributions that comprise this edited volume on historical Ibero-Romance morpho-syntax. Concretely, we will first detail what connections they share and how they differ from each other. Subsequently, we will provide a detailed summary of its contents in order to guide the interested reader, before making some concluding observations.

The common goal of the peer-reviewed contributions of the present volume is to contribute to the field of historical Ibero-Romance morpho-syntax. Several contributions develop fine-grained (micro-comparative) analyses on the basis of historical data ranging from the more widely spoken Ibero-Romance languages, such as Spanish and Portuguese (e.g., Schulte; Eide; Pountain; González Manzano; Moyna) to lesser-studied (but not less interesting) ones, such as Aragonese, Asturian and Catalan (e.g., O'Neill; Pérez Saldanya & Hualde). In the spirit of Maiden's (2004) recommendations for our field to thrive, several papers in this volume have also taken a macro-comparative approach and extend the scope of their investigation to other Romance languages (e.g., Rodríguez Molina & Enrique-Arias) and, of course, their Latin ancestor (e.g., Wright; Elvira; Rodríguez Molina & Enrique-Arias), while yet others highlight parallelisms and differences with other language families (e.g., Elvira; Melis & Flores; Rodríguez Molina & Enrique-Arias). Without a doubt, the biggest asset of the present volume lies in the richness of data that it provides: every single contribution, regardless of the theoretical framework they adhere to or advocate for, sprouts from the unequivocal respect for historical data, which has, in fact, led some authors to the dark corners of insufficiently explored archives and manuscripts in order to unlock the data treasures that lie hidden in unpublished documents (e.g., Calderón Campos; Rodríguez Molina & Enrique-Arias; Torres Cacoullós & Bauman).

An equally important quality of the present volume is the diversity it presents in terms of approaches: cognitive-functional (e.g., Calderón Campos; Melis & Flores; Schulte; González Manzano), generativist (e.g., Eide; Rodríguez Molina & Enrique-Arias), and sociolinguistic and variationist insights (e.g., Pountain, Torres Cacoullós & Bauman) are combined in one book. This diversity stems from the view that divisions between different frameworks are not conducive to a better understanding of language change. Ideally, we need to strive for an open-minded discussion to complement the different strengths that each approach has to offer, reducing as such the chasm that exists between formal and non-formal approaches so that we can reach fully-integrated, ‘holistic’, historical accounts of the morpho-syntactic evolution of Ibero-Romance. The present volume is a step in this direction since – crucially – all contributions, despite the different approaches, have tried to remain accessible to scholars of distinct persuasions and specialisms.

As regards the gamut of phenomena scrutinised in this edited volume, it ranges from morpho-phonological puzzles (e.g., verbal velar allomorphy in O’Neill) and word-formation (e.g., the evolution of concatenative compounds in Moyna) to syntax and interface-related phenomena (e.g., the use of the question particle *si* in Rodríguez Molina & Enrique-Arias; Stylistic Fronting in Elvira; the varying interpretations of subjects in Eide and Schulte; the [definite article + proper noun] construction in Calderón Campos; the development of discourse markers in Pountain and González Manzano; the rise of quirky subjects with the verb *gustar* ‘to like’ in Melis & Flores; the grammaticalisation of *para* in Torres Cacoullós & Bauman and of the concessive subordinators in Pérez Saldanya & Hualde), to, as a coda, methodological suggestions made by Wright for future research in historical Ibero-Romance.

2. Individual contributions

The first contribution by Paul O’Neill deals with the historical development of Ibero-Romance velar allomorphy, i.e. verbal lexemes whose inflectional forms display a velar occlusive before the desinential endings, as in Portuguese *digo* ‘I say’ and Spanish *caigo* ‘I fall’. It is argued that the paradigmatic distributions of velar verbs are “a matter of pure morphology”, as the cells that display velar allomorphy constitute a morpheme, i.e. a semantically heterogeneous collection of paradigm cells. To illustrate, velar allomorphy in Modern Spanish displays the so-called L-pattern as the velar element is present in the 1sg present indicative and in all of the present subjunctive forms. Moreover, morphemes can overlap: the L-pattern is found to overlay with the N-pattern, typically found in verbal paradigms displaying the diphthong-monophthong alternation, as is the case in Old Spanish.

O'Neill convincingly presents evidence from a number of Ibero-Romance varieties that suggests that there can be diachronic and diatopic variation with regards to the morphological parsing of the velar element. The fact that there are no longer verbs in which diphthongisation coincides with velar allomorphy in Modern Spanish, unlike in Old Spanish and its central Ibero-Romance neighbours, is claimed to be due to velar allomorphy no longer being exclusively a function of the endings of verbs. Instead the velar element is viewed as an allomorph of the root. In other words, while in Old Spanish the velar component appears to be parsed as part of the desinential endings, Modern Spanish velar allomorphy is part of the lexical root. Interestingly, data from different central Ibero-Romance varieties, such as Asturian and Aragonese, corroborate that there can, indeed, be variation as to the morphological parsing of the velar occlusive. O'Neill's study confirms thus that morphemes can be correlated with a particular phonological type of allomorphy (velar allomorphy, diphthongised stems or a combination of both), but also with a particular morphological type (root or desinential allomorphy).

The second article in this volume also deals with a morphologic topic, albeit in the sphere of historical word formation. María Irene Moyna examines the evolution of the formal characteristics of concatenative compounds, whose constituents belong to the same lexical category and structural hierarchy, such as *rojinegro* 'red-black' ($[A+A]_A$), *compraventa* 'sale', lit. 'sale-purchase' ($[N+N]_N$) or *ganapi-erde* 'losing game', lit. 'win-lose' ($[V+V]_N$). Comparative discussion of concatenative compounds in Romance, points towards the Latin origin of the linking vowel *-i-*. Moyna further analyses the relative frequency of use of each of the structural patterns used to create concatenative compounds: (i) the structure with overt coordination ($[XconjX]$, *ajo y queso*); (ii) the construction with the linking vowel *-i-* ($[XiX]$, *ajiqueso*); and (iii) the juxtaposition of constituents ($[XX]$, *ajiqueso*). She concludes that, while simple juxtaposition in concatenative compounds has existed since the Middle Ages, over time the pattern with a linking vowel between constituents has become more frequent.

Moyna's study also tries to determine whether a general historical sequence can be proposed for the different patterns of concatenative compounds, and whether empirical data can lend support to the hypothesis whereby overt coordination is at the basis of all concatenative compounds ($[XconjX] > [XiX] > [XX]$). Although the etymological path from the coordination to the linking vowel pattern appears to affect certain adjectival and verbal compounds, it does not appear to hold for nominal compounds. Moreover, the dating of allomorphs appears to indicate the opposite direction of the change, as, in general, the juxtaposed compounds are precursors of equivalent compounds with linking vowels. Instead, it is proposed that linking vowels in concatenative compounds result from the extension of use

of linking elements present in other Spanish compound patterns, such as head-final compounds (e.g. *barbirrapado* ‘beardless’, lit. ‘beard-shaven’ [N+A]_A), in which this vocalic segment first appeared.

A rather understudied structure in the history of Spanish, namely the use of the definite article in front of a proper name (e.g., *la María* lit. ‘the Mary’), is scrutinised with philological care in the contribution by Miguel Calderón Campos. Rather than adopting a syntactic approach, which would treat the definite article as an expletive of sorts, he pursues an analysis on the discourse-pragmatic role of the article before the proper noun. Data for this study is drawn from the Royal Chancery of Granada from 16th to 18th centuries, a virtually unexploited source, as well as CORDE and oral data from contemporary Spanish varieties, such as the one from Granada and various Latin American ones. This paper offers thus an interesting combination of historical and dialectal data, the latter of which are known to constitute a window onto the linguistic past.

Calderón Campos identifies three main discourse-pragmatic functions for the [definite article + proper noun] structure: it can function as (i) a cohesive mechanism for previously mentioned proper names whose presence in the discourse has to be reinforced; (ii) a recognitional element whereby the encyclopaedically shared knowledge is evoked to; and (iii) a marker of informality and colloquiality. The latter is the one currently surviving in various diatopic varieties of Spanish. More precisely, this study uncovers that, while in the contemporary Spanish varieties the presence of a definite determiner endows the noun phrase with the [+informal] marker and signals to the addressee the commonly shared context for the inferences to be activated (“intersubjectification”), in older stages of Spanish the use of this construction also functioned as a marker of discourse prominence or in a recognitional use reserved for the illustrious acclaimed characters of the classical world. It is therefore shown that the discourse-pragmatic conditions that underpinned the rise of the definite article from Latin to Romance are to be found in the use of the definite article before proper noun in Spanish too, subject to diachronic and dialectal variation.

The fourth paper by Javier Elvira analyses the structural and discourse-informational properties of Stylistic Fronting (SF) in Old Spanish, a phenomenon in which a constituent that usually appears postverbally is claimed to have moved to the left of the finite verb. With regards to the properties of SF, it has been claimed that in languages such as Icelandic SF is associated with a subject gap. Elvira, on the contrary, demonstrates that it is perfectly possible to encounter SF with overt subjects in Old Spanish, especially with the copulative verb *ser* ‘to be’. Moreover, it is shown that Old Spanish SF is attested both in main and subordinate clauses, even though the motivations behind this phenomenon seem to be different in each context.

Although an attempt is made to verify whether SF cases have a focal interpretation, the evidence is not conclusive. Instead, the author presents empirical data for syntactic triggers for the appearance of SF, while not excluding possible pragmatic cues. Despite the existence of V1 in Old Spanish, it is claimed that in main clauses SF appears to be caused by the absence of an element in the preverbal domain, especially as copulative or auxiliary verbs like *ser* 'to be' and *aver* 'to have' systematically do not appear in initial or post-topical position. The distribution of SF with an unstressed pronoun is then compared to the occurrence of analytic future constructions, which are seen as the result of the interference of enclisis of the unstressed pronouns. SF in subordinate clauses is shown to be conditioned in a different way since negative adverbs do not block its occurrence, unlike in main clauses. This is taken to suggest that the preverbal position is not essential to understand the movement of elements of the verbal phrase in subordinate clauses and that this word order pattern can be understood as a residue of the old Latin verb-final order.

The following paper by Javier Rodríguez Molina and Andrés Enrique-Arias examines the particle *si* and in particular how it functions in polar questions. Although *si* in polar questions is carefully disentangled from both the use in conditionals and polar indirect (or embedded) interrogatives, there is still a diachronic connection between the three.

Classical Latin had a rich system of interrogative particles, which in its transition to Late Latin underwent a series of substantial changes. Particularly relevant to the phenomenon under study is the replacement of NUM and NUMQUID by *si* in indirect questions, a structure that would eventually prevail in all Medieval Romance languages. From this stage onwards there seems to be two different evolution paths for *si* which mutually reinforced each other: (i) internal conditioning which saw the expansion of *si* from indirect interrogatives to polar (direct) ones, such as the semantic contiguity between certain conditional and interrogative contexts; and (ii) exogenous change since *si* was adopted in the Latin used by Bible translators and Christian writers to render εἰ in the Greek original and Hebrew particles *ha* and *im* in indirect questions.

In particular, the historical trajectory, which is claimed to be common to all Romance languages, essentially entails a grammaticalisation pathway from conditional > indirect interrogative > polar question marker. The mechanics of such reanalysis entails a change from a complex structure, i.e. an embedded question, to a simple one. This phenomenon is regarded as a case of insubordination. This analysis seems to accurately capture how the particle *si* derives from a subordinate structure (the Latin indirect interrogative), which has become a main clause structure via ellipsis, before dying out after 17th century.

Many languages, like Russian, among others, have constructions whereby the subject shows a dative or oblique case, i.e. quirky subjects. Chantal Melis and

Marcela Flores analyse the Spanish verb *gustar* ‘to like’, which switched from a nominative to a dative subject/experiencer. The fundamental objective of the study is to provide some insight into the semantic and pragmatic factors that motivated this syntactic shift.

From the Medieval Spanish period onwards, *gustar* went through three main stages: (i) from the 11th to the 15th century, *gustar* had the meaning of a taste activity verb; during this stage, the experiencer of the verb was assigned nominative case, while the thing which was tasted functioned as the direct object; (ii) from the 16th to most of the 18th century, *gustar* became a mental verb, joining the category of volitional “liking” verbs; during this phase, the human participant continued to function as the subject or experiencer of the predicate rather than the physical one; (iii) from 18th century onwards we observe competing grammars: although a dative experiencer for *gustar* became possible, the nominative was still preferred. This variation phase shows interference with the syntax of *placer* ‘to please’, which always selected for a dative experiencer. Over time the dative experiencer became more frequent and replaced almost entirely the nominative one.

To highlight the role of semantics in the development of *gustar*, near-synonyms are considered: *amar* ‘to love’, *querer* ‘to want’ and *placer* ‘to please’. The important contribution of this analysis lies in motivating semantic factors for what is typically viewed as syntactic change. For instance, a context in which the stimulus participant of the verb corresponds to the discourse topic favours the use of a dative construction. Furthermore, the *gustar* construction assigns subject characteristics to the participant, such as verbal control agreement, as well as a sense of responsibility for the pleasure felt by the experiencer. Thus, what this shift reflects is essentially a semantic change of perspective: from the experiencer who takes pleasure (nominative construction) to the stimulus that causes pleasure to the experiencer (dative construction). From a pragmatic perspective, the study reveals that in a context in which speakers aim to communicate something about their mental state, they tend to select the dative construction rather than the nominative one.

The following contribution by Kristine Gunn Eide studies the changes that affected the subject position in the history of Portuguese; in particular, the change from a verb-subject (VS) to a subject-verb (SV) order. Eide relies on certain well-established assumptions about Modern European Portuguese: (i) in unmarked contexts (as a reply to “what happened?”), both SV and VS orders are allowed, while (ii) old information subjects show a SV pattern; and (iii) new information S show a VS pattern. As such, the article explores the interplay between syntax and Information Structure of the subject position in order to account for the consolidation of SV constructions with intransitive verbs in the history of Portuguese.

Based on an empirical corpus with data from the 13th to the 20th century, the author proposes two distinctive patterns: while in Old Portuguese there is a trend towards VS with intransitive verbs, in Classical Portuguese and Modern European Portuguese the placement of the subject is regulated by the discourse value (old vs. new): (i) when the subject has already been mentioned (old), it shows the same order as with transitive verbs namely, SV[O]; and (ii) when the subject contains new information, SV seems to be predominant.

Statistical evidence from the corpus leads to a more nuanced picture whereby SV depends both on the nature of the predicate and the information structure of the subject. More explicitly: (i) with transitive and unergative verbs in Old and Classical Portuguese it is possible to find both VS and SV orders, independently from the information structure load of the subject; nevertheless, in the transition from the Classical to Modern period, VS constructions decrease when the subject is old information; (ii) in the older periods, in transitive verbs the position of the object in SVO orders may depend on the presence of adverbs. In modern times, SVO tends to be the norm with transitive verbs; (iii) with intransitive verbs, the corpus shows two changes in the transition from Old to Modern European Portuguese: the progressive decrease of VS structures with subjects with old information, and the increase of SV structures when subjects convey new information.

The seventh paper, by Kim Schulte, aims to identify contrastively whether the distribution of overt subjects in Spanish and Portuguese infinitives is similar, both from a diachronic and a synchronic point of view. Corpus data from 12th and the 20th century reaffirm that all types of infinitival clauses may contain an overt subject in either language, be it a subject pronoun or a noun phrase, at some historical point. A statistical comparison between Spanish and Portuguese demonstrates that 33% of Portuguese infinitives have an overt subject, in contrast to only 1% of Spanish infinitives. It is worth noting that in Portuguese the morphological agreement between the infinitive and the overt subject is compulsory for the second person singular and all persons in plural. Furthermore, it is claimed that, contrary to what some grammarians have suggested, most overt subjects in Spanish infinitival clauses are co-referential to those of the main clause. Additionally, while similar types of clauses in Spanish and Portuguese can contain overt subjects, the subject position vis-à-vis the infinitive may differ.

In Old Portuguese infinitive constructions overt subjects frequently appear preverbally partly due to the default SVO word order. In Classical Portuguese, there is a tendency for pronominal subjects to appear preverbally, while subject DPs occur postverbally. This change in the position of subject DPs, which unlike pronominal subjects usually provide new information, is likely to be due to a generalised word order shift during said period since it is noted that subjects in finite clauses also appear mostly in a postverbal position. Modern Portuguese witnesses

a shift back to the SVO word order whereby the preverbal position for the infinitival subject is only overridden by an overt intensifier (triggering contrastive focus) and thus resulting into postverbal subjects.

At the early stages of Spanish we also find preverbal subjects with infinitives, but, unlike Portuguese, this pattern did not become entrenched in Spanish. Overt infinitival subjects had two functions: (i) disambiguation; and (ii) contrastive focus. Contemporaneous data, contrary to general belief according to which the subject in peninsular personal infinitives is always postverbal, reveal that the preposition that introduces an infinitival adverbial clause can influence the position of the overt subject in relation to the infinitive. Another influencing factor for the position of the overt subject is the presence of clitic object pronoun in combination with certain prepositions such as *para*, *antes de* and *después*, all of which increase the chances of a postverbal subject.

Rena Torres Cacoullous and Joseph Bauman examine the grammaticalisation of the preposition *para* in the history of Spanish. Such a study is apt for big scale data analysis and this is precisely what the authors deliver. The study of the evolution of *para* also calls for a meticulous study of the manuscripts to resolve cases of graphical ambiguity – again on this front the authors deliver and show unequivocal attestations of *para* dating from as early as the 13th century.

Their findings show that the earliest texts show non-agglutinated forms, *por a* and *por al*, as well as agglutinated ones in almost identical contexts, the latter forms persisting as late as the 17th century. Although two-preposition combinations are common in Old Spanish, there are by no means uniquely found in Spanish, since most Romance varieties show contractions/combinations (e.g., Galo- and Italo-Romance). Yet, *para* is uniquely found in Ibero-Romance. The careful examination of the manuscripts shows that the orthographic variants *por a*, *póra*, and *para* were used contemporaneously and that the process of the fusion of the two structural elements was gradual, as is expected.

The authors study the kinds of object NPs with which *para* co-occurs diachronically and they discover that it combines with motion verbs whereby the NPs that denote the endpoint of the motion are a location (spatial allative use); or may even be a person. Therefore, during these early stages the semantic contribution of the preposition *a* was still discernible. Crucially, the main change to be observed concerns the spatial uses of the preposition, which steeply decline from the 14th and 15th centuries onwards. A follow-up change, observed in the 18th century, is that *para* becomes more associated with the purposive infinitive to the detriment of *por*, although early stages of co-variation show a slight meaning differentiation.

Torres Cacoullous and Bauman demonstrate that, considering that *para* started as a sequence of two prepositions, chunking took place due to the repeated and

frequent use of the co-occurring components of this preposition. Chunking conditions aside, this type of allative-to-purposive evolution is cross-linguistically attested (e.g., Lezgian).

The article by Manuel Pérez Saldanya and José Ignacio Hualde scrutinises the evolution of concessive conjunctions in Spanish and Catalan, which are cross-linguistically susceptible to change. In particular, they analyse two distinct phenomena: (i) the evolution of *aunque/encara que* from conditional concessive meaning ('even if') to concessive ('although') and adversative ('but'); and (ii) the evolution of *puesto que/posat que* from concessive to causal, which represents an exception to the standard tendency of change from a causal to a concessive function.

The origin of Spanish and Catalan concessives is to be found in causal and conditional constructions and in constructions with quantifiers (*por más que/por mucho que*) that express intention, desire, will or fear. *Aunque*, which replaces the older *maguer que*, derives from the scalar adverb *aun* (< ADHUC 'until now') which modifies a subordinate clause starting with *que*. The unification into *aunque* signals a greater degree of grammaticalisation. It is first attested in the second half of the 13th century with conditional value. However, it gradually drifted to a concessive function, potentially already implicit in the scalar value of *aun*, which becomes the prevailing concessive conjunction by the 17th century. The Catalan counterpart *encara que* followed the same diachronic path, but at a slower rate.

The second pair of concessive constructions the authors study is the exception presented by *puesto que/posat que* to the standard path of change from causal to concessive meaning. These conjunctions, in fact, follow the opposite trend. Pérez Saldanya and Hualde argue that such a change has to be understood as an example of reinterpretation that was made possible by the polysemy exhibited by absolute constructions in which *puesto* can convey two meanings 'supposed' and 'set down', the latter of which prevailed and stabilised in its modern uses.

The next contribution by Mónica González Manzano deals with the diachronic development of the Spanish epistemic adverbs *verdaderamente* 'truly', *realmente* 'really', and *ciertamente* 'certainly', which grammaticalised from verbal adverbs to sentence adverbs to discourse markers. To start with, the author traces this grammaticalisation (or pragmaticalisation) process, which, as is usual, manifests itself through various micro-changes at different levels of the language. Syntactically, she observes a shift towards the left periphery, as these adverbs start appearing preverbally, widen in structural scope, and gain in autonomy. Semantically, the meanings of *verdaderamente*, *realmente* and *ciertamente* become gradually more subjective and start conveying pragmatic nuances, albeit without losing their original meanings. As such, these expressions exhibit a layering of various meanings.

Additionally, a second grammaticalisation process is discussed, whereby a new intra-sentential adverbial function emerged, to wit, the ability to modify adjectives with an intensifying meaning. It is claimed that the locus of this change, i.e. the critical contexts that are ambiguous and thus allow a new function to surface, must have been cases in which these elements modify past participles, due to their hybrid formal characterisation and to the fact that these contexts also allow an intensifying reading. The extension of this new function is then sketched as a process that affects first prototypical adjectival contexts, such as scalar adjectives, and second less prototypical cases, such as superlative and non-scalar adjectives. In sum, the abovementioned adverbs are shown to be (come) highly polyfunctional and multi-layered as they experience a meaning extension. Given that the original meaning of these adverbs shifts into a more pragmatic interpretation, they eventually turn into a discourse marker, while also developing an intensifying interpretation as an adjective modifier.

Christopher J. Pountain describes in his contribution the frequency, distribution, and functions of an array of discourse markers found in four mid-16th century Spanish texts. The starting point for this study are literary texts by Lope de Rueda, which are contrasted with semi-literary and non-literary texts from the contemporaneous Tomás de Mercado, Santa Teresa, and Fray Luis de Granada. The examined discourse markers comprise (i) the commentary marker *pues* ‘well, in that case’; (ii) a variety of ordering markers, which include opening, continuity and closure markers, such as *lo primero* ~ *primeramente* ‘first’, *últimamente* ‘finally’, but also markers like *mayormente* ‘more importantly’ and *especialmente* ‘particularly’, and, finally; (iii) consecutive markers, such as *así* ‘so then’.

Various variables are shown to contribute to the differing uses and frequency profiles of the analysed discourse markers in 16th-century Spanish. Concretely, the differences in the presence/absence and frequencies of use of the analysed discourse markers appear to be dependent on:

1. the dimensions of text type in which the discourse markers appear, i.e. the tenor and field of the text: for instance, contrastive markers, such as *por una parte* [...], *por otra* [...] ‘on the one hand [...], on the other [...]’ appear to be characteristic of dialectic and didactic prose;
2. the register of the text: e.g., the commentary function of *pues* ‘well, in that case’ is used very abundantly in the dialogues of Rueda’s plays, while it is hardly ever attested in the formal register of Tomás de Mercado’s *Summa de tratos*, which is a treatise on business ethics;
3. the (intended representation of) social class: for instance, discourse markers of order are associated with cultured speakers; and, finally,

4. stylistic and idiolectal preferences of the authors, as is observed by the lexical variation in the use of the near-synonyms *primeramente* ~ *lo primero* ‘first’, the former of which is favoured by Fray Luis de Granada and the latter by Tomás de Mercado.

In sum, Pountain presents compelling arguments that demonstrate that the variation in 16th-century Spanish discourse markers is determined not only by text, discourse, and author dependent factors, but, importantly, by social class as well; thus, discourse markers with variant exponents for the same function can eventually be interpreted as social markers or indicators.

In the final chapter of this edited volume, Roger Wright urges linguists who wish to study the semantic, morphological or syntactic development of Old Ibero-Romance phenomena to include, in their investigations, data from before 1200 and, as such, to stop ignoring the wealth of early (Ibero-)Romance textual documentation that exists, but remains almost untouched. Scholars are recommended not only to refer to practices in standard Latin and examine textual material from after the so-called “Twelfth-century Renaissance”, but also to adduce relevant evidence from texts of the intervening thousand years, as there is a huge amount of potentially relevant data available for the period in between. Wright reminds us that it is a quantitative fact that there survived more written Latin material from after the end of the Roman Empire, much of it ecclesiastical in nature admittedly, than from all other periods. This is the case for texts proceeding from areas where Latin was a foreign language, but also where the native speech was a kind of Late Latin, i.e. Early Romance. The history of Latin in the Iberian Peninsula before 1200 was similar to that of any other literate living language: although the writers of the Early (Ibero-)Romance texts might have been imitating earlier models and/or were using a high style, their writing and speech can be considered as two modes of the same language. In view of this, written records can also attest to the evolved Early (Ibero-)Romance features and thus they can be used by historical linguists, with care always, as material for the study of the spoken language of their time.

3. Coda

In conclusion, we would like to remind the reader that, ultimately, whatever approach or framework be chosen, or whatever language or phenomenon be examined, unequivocal respect for (historical) data should always be at the heart of the (historical) linguist’s enterprise.

Bearing the previous in mind, we cannot think of a better way to conclude this introductory chapter other than echoing Roger Wright’s methodological concerns:

while doing so, we should not shy away from studying lesser known old (Ibero-) Romance sources, that is the wealth of data prior to 1200. But this is something to aspire to for future historical (Ibero-)Romance research.

Reference

Maiden, M. (2004). A necessary discipline: Historical Romance linguistics. *La corónica*, 32(2), 215–221. doi:10.1353/cor.2004.0036

Velar allomorphy in Ibero-Romance

Roots, endings and clashes of morphemes

Paul O'Neill

University of Sheffield

This article analyses the historical development of velar allomorphy in the Ibero-Romance verb and argues that the recognition of (i) autonomous morphological structures, or morphemes (Aronoff, 1994), and (ii) variation as to the morphological parsing of velar allomorphy can facilitate more convincing explanations of the evolution, spread and change of this allomorphy and the class of velar verbs which exhibit it. With regards morphemes, this study presents data, which prompts the conclusion that the allomorphy in these cells tends not only to converge upon one particular phonological form (Maiden, 2004), but also upon a particular morphological structure or type. Regarding the morphological parsing of the allomorphy, it is argued that it can correspond to root allomorphy or desinential allomorphy.

Keywords: Ibero-Romance; verb; velar allomorphy; morpheme; Spanish; Portuguese; Aragonese; Asturian

1. Introduction

Velar allomorphy refers to the inflectional forms of a lexeme which display a velar occlusive consonantal phoneme, either voiced /g/ or voiceless /k/, before the desinential endings of person and number and/or tense and mood; this allomorphy displays a particular paradigmatic patterning. The classical example of such velar allomorphy is the Spanish verb *caer* 'fall' in which the velar element occurs exclusively in the 1SG present indicative and all of the present subjunctive. I follow Maiden (2004) and refer to this particular paradigmatic patterning as

the L-pattern, as shown in Table 1.¹ The verbs which display this allomorphy are termed velar verbs.

Table 1. The L-pattern morpheme

	Pres. Indic.	Pres. Subj.	Future	Conditional
1SG	<i>caigo</i>	<i>caiga</i>	<i>caeré</i>	<i>caería</i>
2SG	<i>caes</i>	<i>caigas</i>	<i>caerás</i>	<i>caerías</i>
3SG	<i>cae</i>	<i>caiga</i>	<i>caerá</i>	<i>caería</i>
1PL	<i>caemos</i>	<i>caigamos</i>	<i>caeremos</i>	<i>caeríamos</i>
2PL	<i>caéis</i>	<i>caigáis</i>	<i>caeréis</i>	<i>caeríais</i>
3PL	<i>caen</i>	<i>caigan</i>	<i>caerán</i>	<i>caerían</i>
	Imp. Subj. (-ra)	Imp. Subj. (-se)	Preterite	Imp. Indic.
1SG	<i>cayera</i>	<i>cayese</i>	<i>caí</i>	<i>caía</i>
2SG	<i>cayeras</i>	<i>cayeses</i>	<i>caíste</i>	<i>caías</i>
3SG	<i>cayera</i>	<i>cayese</i>	<i>cayó</i>	<i>caía</i>
1PL	<i>cayéramos</i>	<i>cayésemos</i>	<i>caímos</i>	<i>caíamos</i>
2PL	<i>cayerais</i>	<i>cayeseis</i>	<i>caísteis</i>	<i>caíais</i>
3PL	<i>cayeran</i>	<i>cayesen</i>	<i>cayeron</i>	<i>caían</i>
	Imperative	Infinitive	Gerund	Participle
2SG	<i>cae</i>	<i>caer</i>	<i>cayendo</i>	<i>caído</i>
2PL	<i>caed</i>			

Velar allomorphy has been of special interest to Romance linguists since, with the exception of French, Rhaeto-Romance and the majority of Sardinian varieties, this velar allomorphy is attested in all other Romance languages. The different languages, however, vary as to the paradigmatic distribution of this allomorphy and the number of lexemes which display such allomorphy.

Ibero-Romance is of particular interest in this respect since Portuguese displays only one velar verb, *dizer* 'say', whose allomorphy and paradigmatic distribution (the L-pattern) is entirely predictable on the basis of regular sound change. Modern Spanish contains approximately 155 velar verbal roots, excluding their

1. The terms 'L-pattern' and 'N-pattern' have no semantic or phonological significance. They were adopted by Maiden (2004) because in conventional representations of the verb paradigm on paper, the cells which comprise the L-pattern resemble a rotated letter 'L' while those in the N-pattern look like the letter 'N' in Morse code.

derivatives, which display a voiceless velar allomorph² in the L-pattern and 11 verbal roots, again excluding derivatives, with a voiced velar allomorph.³ These verbs have been of interest to historical linguists since most velar allomorphy is non-etymological. In fact, of the voiced velar allomorphy the only lexeme whose allomorphy is etymologically expected is *decir* ‘say’.

Catalan, along with Occitan, are the most bountiful languages with regards velar verbs since such verbs tend to be some of the most frequently used verbs in the language; according to Wheeler’s calculations (2011, p. 184), for the Barcelona standard variety of Catalan there are approximately 80 verb roots (316 lexemes) which display velar allomorphy, voiced and voiceless.⁴ In Catalan the distribution of this allomorphy extends beyond the L-pattern cells and into the synthetic preterite, the imperfect subjunctive and the gerund; this is illustrated in Table 2 for the verb *moure* ‘move’.

Additionally, there are also 19 or 20 roots (101 to 113 lexemes taking into account derived forms) which display velar allomorphy in all the cells above with the exception of the past participle in which there occurs an irregular rhizotonic participle.⁵ According to Wheeler (2011, p. 184), there are no verbs in Modern Catalan in which the velar allomorph in its paradigmatic distribution is expected etymologically.

Within the different languages, there is also much diachronic variation as to which lexemes displayed velar allomorphy. Thus, in Old Spanish there was a higher

2. The full list of these L-pattern verbs can be found in the Appendix. The majority of these verbs are denominal or deadjectival verbalisations e.g. *humedecer* ‘make humid’, *entontecer* ‘make stupid’, *anochecer* ‘become dark/nighttime’. This list of verbs also includes some frequent verbs not of this type, e.g. *apetecer* ‘feel like’, *parecer* ‘seem’ and derivatives thereof *aparecer* ‘appear’, etc.

3. These verbal roots include *decir* ‘say’, *hacer* ‘do’, *salir* ‘go out’, *valer* ‘be worth’, *poner* ‘put’, *venir* ‘come’, *tener* ‘have’, *caer* ‘fall’, *traer* ‘bring’, *roer* ‘gnaw’, *oír* ‘hear’ plus their derivatives.

4. Of these 80 roots, six roots vary as to whether they display the velar allomorph in all the cells; four roots display it only in the 1SG present indicative and all the forms of the present subjunctive and seven in the synthetic preterite, imperfect subjunctive and past participle only. Note also that the verb *córrer* ‘run’ in the Barcelona standard, never exhibits a velar in the 1SG present indicative and the rhizotonic forms of the present subjunctive but can display one variably in the 1PL and 2PL present subjunctive only. In the other synthetic preterite, the imperfect subjunctive and the past participle, however, this verb consistently displays velar allomorphy. The verb *escriure* ‘write’ consistently inflects with the velar consonant in the L-pattern cells, but in the other cells there is variation as to whether it possesses the velar allomorph or not. This, however, is not the case with the past participle in which the velar does not occur.

5. Cf. *prendre* ‘take’, 1SG present indicative *prenc*, 1SG present subjunctive *prengui*, 1SG preterite *prenguí*, 1SG imperfect subjunctive *prenguéis*, past participle *pres*.

Table 2. Velar allomorphy in *moure*

	Pres. Indic.	Pres. Subj.	Future	Conditional
1SG	<i>moc</i>	<i>mogui</i>	<i>mouré</i>	<i>mouria</i>
2SG	<i>mous</i>	<i>moguis</i>	<i>mouràs</i>	<i>mouries</i>
3SG	<i>mou</i>	<i>mogui</i>	<i>mourà</i>	<i>mouria</i>
1PL	<i>movem</i>	<i>moguem</i>	<i>mourem</i>	<i>mouríem</i>
2PL	<i>moveu</i>	<i>mogueu</i>	<i>moureu</i>	<i>mouríeu</i>
3PL	<i>mouen</i>	<i>moguín</i>	<i>mouran</i>	<i>mourien</i>
	Imp. Indic.	Imp. Subj. (-se)	Preterite	
1SG	<i>movie</i>	<i>mogués</i>	<i>mogui</i>	
2SG	<i>movies</i>	<i>moguëssis</i>	<i>mogueres</i>	
3SG	<i>movie</i>	<i>mogués</i>	<i>mogué</i>	
1PL	<i>movíem</i>	<i>moguëssim</i>	<i>moguérem</i>	
2PL	<i>movíeu</i>	<i>moguëssiú</i>	<i>moguéreu</i>	
3PL	<i>movien</i>	<i>moguëssin</i>	<i>mogueren</i>	
	Imperative	Participle	Gerund	Infinitive
2SG	<i>mou</i>	<i>mogut</i>	<i>movent</i>	<i>moure</i>
2PL	<i>moveu</i>			

proportion of etymological velar verbs e.g., TANGŌ > *tango* 'I play', FRANGŌ > *frango* 'I break', CINGŌ > *cingo* 'I surround', TINGŌ > *tengo* 'I dye', SPARGŌ > *espargo* 'I scatter', RINGŌ(R) > *ringo* 'I scold'. The velar allomorphy in these verbs, however, was eradicated whilst, paradoxically, it was extended to other verbs in which it was not etymologically expected: *salgo*, *valgo*, *pongo*, *vengo*, *firgo*, *caigo*, *traigo*, *oigo* (which all persist in Modern Spanish) and *suelgo*, *muelgo*, *duelga*, *calga*, *fuigo*, *fiergo*, *remango*, *tuelgo*, *distruigo*, *restituigo*, *falgo* < *fallir* (which are only attested in Old Spanish).

The linguistic factors which determine the distribution of velar allomorphy in Romance have been a contentious matter for linguistic theory since some have proposed that the allomorphy is a matter of phonology (St. Clair & Park, 1974; Bermúdez-Otero, ms. for Spanish; Burzio, 2004 for Italian; Hualde, 1992, p. 410; Mascaró, 1983 for Catalan) others have claimed that it is a matter of pure morphology (Maiden, 2001, 2004; O'Neill, 2011a, pp. 204–243, 2015 for Spanish; Maiden, 2001, 2004, 2009; Pirrelli, 2000, pp. 79f., 178–184; Pirrelli & Battista, 2000 for Italian; Wheeler, 2011 for Catalan), and others (Bybee, 1985 for Spanish; Pérez-Saldanya, 1995 for Catalan) have highlighted the importance of semantic notions

of markedness and explained the paradigmatic distribution by recourse to the 1SG present indicative being classed as unmarked and forming the base upon which all the other more marked velar forms are derived. Moreover, both within and across the different theories as to what grammatical component is responsible for the distribution of the velar element, there is no consensus as to the morphological parsing of the velar consonant: some consider it to constitute part of the allomorphic root or stem (Maiden, 2001, 2004 for Spanish; Roca Pons, 1968, p. 231; Mascaró, 1983, pp. 155–194; Badia, 1994, pp. 574–578 for Catalan), others an infix (Malkiel, 1974 for Spanish; Perea, 2002, pp. 597–601; Pérez-Saldanya, 1998, pp. 73–75; Viaplana, 1984; Wheeler, 1993, p. 196 for Catalan), and others part of the desinential endings (Martín Vegas, 2007; p. 168; O'Neill, 2011a, pp. 216–220, 2015 for Spanish).

In this article I draw upon previous studies (Maiden, 2001, 2004, 2009; O'Neill, 2011a, 2015) and I argue that velar allomorphy in Ibero-Romance is morphologically determined; more specifically, I contend that the cells which display velar allomorphy constitute a morpheme (Aronoff, 1994). With regards the morpheme and velar allomorphy, I present evidence from a number of Ibero-Romance varieties which suggest that there can be diachronic and diatopic variation as to the parsing of the velar element. At times, it can be considered to be part of the lexical root, other times the evidence suggests it has been parsed as part of the desinence. I argue that the recognition of such purely morphological structures (morphemes), coupled with an appreciation of the variation in the morphological parsing of the velar element offers a way to understand the diatopic and diachronic variation of the lexemes which display/displayed velar allomorphy in Ibero-Romance.

The organisation of this article is the following. In Section 2, I briefly describe the origins of velar allomorphy in Ibero-Romance. In Section 3, I summarise the arguments for believing that the velar element forms part of the desinential endings in Old Spanish and that it is morphologically conditioned. In Section 4, I add to these arguments but present evidence which prompts the conclusion that in further developments in Spanish the velar element was analysed as part of the root. In Section 5, I introduce another morpheme, the N-pattern morpheme, and I present data from central Ibero-Romance varieties of how the L-pattern and N-pattern morphemes interact in support of the hypothesis that there can be variation and change as to the parsing of the velar element and that morphemes can converge on different morphological structures and phonological exponences. In Section 6, I advance an argument as to why an appreciation of the ideas developed in this article can help explain the diachronic and cross-linguistic variation regarding the lexemes which form the velar class.

2. Brief explanation as to the origin of velar allomorphy in Ibero-Romance

The source of velar allomorphy is phonological since it is due to the absence of palatalisation of velar consonants before front vowels in opposition to its presence before back vowels. In the transition from Latin to Romance, velar consonants in the Iberian Peninsula underwent a process of palatalisation and affrication exclusively before front vowels, which caused a phonemic split of both /k/ and /g/. In Old Spanish, before back vowels these consonants remained velar (although /k/ was subject to voicing and /g/ to becoming an approximant or to being lost).⁶ Before front vowels velar occlusives were supposedly first palatalised and then pronounced as the affricates [dʒ] and [tʃ], depending on the voicing contexts (VĪCĪNU > *vezino* 'neighbour' vs. LOCĀLE > *logar* 'place', Penny, 2002, p. 66).

Within the verb morphology, the verbs which met the criteria to be sensitive to these different phonetic contexts preserved their root-final velar consonants exclusively in the L-pattern cells since the desinences of these cells contained back vowels.⁷ The phonological outcome of the root-final velar consonant in the non-L-pattern cells was not uniform but depended upon the phonetic context. When the velar consonant was preceded by a vowel (or by a rhotic consonant), it was subject to the normal process of affrication and voicing (see Table 3 and Table 4).

Table 3. Outcome of DICŌ 'say', REDUCŌ 'reduce' and ADDUCŌ 'adduce'^{*}

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>digo</i>	<i>diga</i>	<i>redugo</i>	<i>reduga</i>	<i>adugo</i>	<i>aduga</i>
2SG	<i>dizes</i>	<i>digas</i>	<i>reduzes</i>	<i>redugas</i>	<i>aduzes</i>	<i>adugas</i>
3SG	<i>diz(e)</i>	<i>diga</i>	<i>reduz(e)</i>	<i>reduga</i>	<i>aduz(e)</i>	<i>aduga</i>
1PL	<i>dezimos</i>	<i>digamos</i>	<i>reduzemos</i>	<i>redugamos</i>	<i>aduzemos</i>	<i>adugamos</i>
2PL	<i>dezides</i>	<i>digades</i>	<i>reduzedes</i>	<i>redugredes</i>	<i>aduzedes</i>	<i>adugades</i>
3PL	<i>dizen</i>	<i>digan</i>	<i>reduzen</i>	<i>redugan</i>	<i>aduzen</i>	<i>adugan</i>

* Other Latin verbs which could have shown this development were: CŌQ(U)Ō 'cook'.

6. Cf. NEGĀRE > *negar* 'deny', LIGĀRE > *liar* 'tie' (Penny, 2002, p. 68).

7. These verbs were those that (a) were not of the Latin 1st conjugation, (b) did not contain etymological [j] and (c) whose roots ended in a velar consonant.

Table 4. Outcome of SPARGŌ ‘scatter’ and ĘRIGŌ ‘erect’

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>espargo</i>	<i>esparga</i>	<i>yergo</i>	<i>yerga</i>
2SG	<i>esparzes</i>	<i>espargas</i>	<i>yerzes</i>	<i>yergas</i>
3SG	<i>esparz(e)</i>	<i>esparga</i>	<i>yerz(e)</i>	<i>yerga</i>
1PL	<i>esparzemos</i>	<i>espargamos</i>	<i>yerzemos</i>	<i>yergamos</i>
2PL	<i>esparzedes</i>	<i>espargades</i>	<i>yerzedes</i>	<i>yergades</i>
3PL	<i>esparzen</i>	<i>espargan</i>	<i>yerzen</i>	<i>yergan</i>

When the velar consonant was preceded by a nasal consonant the outcome was /ɲ/ (see Table 5).⁸ When preceded by a voiceless consonant (which are principally those verbs which were formed with the Latin ingressive infix -SC- /-sk-/) the outcome was the voiceless affricate /tʃ/ (see Table 6; Penny, 2002, p. 178f.).

Table 5. Outcome of TANGŌ ‘touch’, CINGŌ ‘surround’ and FRANGŌ ‘break’

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>tango</i>	<i>tanga</i>	<i>cingo</i>	<i>cinga</i>	<i>frango</i>	<i>franga</i>
2SG	<i>tañes</i>	<i>tangas</i>	<i>ciñes</i>	<i>cingas</i>	<i>frañes</i>	<i>frangas</i>
3SG	<i>tañe</i>	<i>tanga</i>	<i>ciñe</i>	<i>cinga</i>	<i>frañe</i>	<i>franga</i>
1PL	<i>tañemos</i>	<i>tangamos</i>	<i>ciñemos</i>	<i>cingamos</i>	<i>frañemos</i>	<i>frangamos</i>
2PL	<i>tañedes</i>	<i>tangades</i>	<i>ciñedes</i>	<i>cingades</i>	<i>frañedes</i>	<i>frangades</i>
3PL	<i>tañen</i>	<i>tangan</i>	<i>ciñen</i>	<i>cingan</i>	<i>frañen</i>	<i>frangan</i>

8. Other Latin verbs which could have shown this development are PINGŌ ‘paint’, FINGŌ ‘make up, touch’, *RINGŌ ‘snarl’, STRINGŌ ‘draw tight’, TINGŌ ‘wet, dye’. Also, attested in Old Spanish with the velar allomorph are the verbs FRANGŌ > *frango* ‘I break’ and IUNGO > *ungo* ‘I join, yoke’. As Martín Vegas (2007, pp. 170–171) has pointed out, in Old Castilian, these latter two contained competing allomorphs in the non-L-pattern cells: those ending in a palatal nasal and also ones ending in /ndz/; thus the forms *uñe*, *uñes* from the verb *uñir* ‘yoke’ and *frañe*, *frañes*, *frañir* ‘break’ alternated with forms *unze*, *unzes*, *unzir* and *franze*, *franzas*, *franzir* respectively. This is due to the various results of the Latin consonantal sequence NG + front vowel, which could either result in /ɲ/ (cf. UNGULA > *uña* ‘nail’), /ndz/ (cf. *RINGELLA > *renzilla* ‘quarrel’) and in the case of QUINGĚNTI > *quinientos* ‘five hundred’ the result can even be /n/. For the verb *uñir*/*unzir* the most frequent forms were those ending in the sound /ndz/ from which the modern verb *uncir* ‘yoke’.

Table 6. Outcome of COGNŌSCŌ ‘know’ and *MERESKO ‘deserve’

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>conosco</i>	<i>conosca</i>	<i>meresco</i>	<i>meresca</i>
2SG	<i>cono(s)çes</i>	<i>conoscas</i>	<i>mere(s)çes</i>	<i>merescas</i>
3SG	<i>cono(s)çe</i>	<i>conosca</i>	<i>mere(s)çe</i>	<i>meresca</i>
1PL	<i>cono(s)çemos</i>	<i>conoscamos</i>	<i>mere(s)çemos</i>	<i>merescamos</i>
2PL	<i>cono(s)çedes</i>	<i>conoscades</i>	<i>mere(s)çedes</i>	<i>merescades</i>
3PL	<i>cono(s)çen</i>	<i>conoscan</i>	<i>mere(s)çen</i>	<i>merescan</i>

This velar allomorphy and its particular paradigmatic patterning underwent two opposing types of development. Firstly, the velar allomorphy was eradicated from those verbs in which it was etymological. This development is attested in Old Spanish for a number of verbs in which the most frequently occurring form, usually 3SG present indicative, replaces the allomorphy in the L-pattern e.g., TANGŌ > *tango* >> *taño* ‘I play’, FRANGŌ > *frango* >> *fraño* ‘I break’, CINGŌ > *cingo* >> *ciño* ‘I surround’, TINGŌ > *tengo* >> *tiño* ‘I dye’, SPARGŌ > *espargo* >> *esparzo* ‘I scatter’, RINGŌ(R) > *ringo* >> *riño* ‘I scold’.⁹ The motivations behind this loss of velar allomorphy will be discussed at length in Section 4. Secondly, the velar allomorphy was extended to other verbs in which it was not etymologically expected (*salgo, valgo, pongo, vengo, firgo, caigo, traigo, oigo*, which all persist in Modern Spanish, and *suelgo, muelgo, duelga, calga, fuigo, fiergo, remango, tuelgo, distruigo, restituigo, falgo* < *fallir*, which are only attested in Old Spanish).

It should be noted that the overwhelming majority of verbs which displayed non-etymological velar allomorphy are those for which, given the effects of desinential *yod* in the verbal paradigm (O’Neill, 2012), ought to have shown a particular type of allomorphy, usually characterised by a palatal consonant, exclusively in the L-pattern cells. Examples of this allomorphy are given in Table 7 for the present tense paradigms of proto Ibero-Romance reflexes of FACERE ‘do’, TENĒRE ‘have’, SALĪRE ‘go out’, AUDĪRE ‘hear’, UENĪRE ‘come’, UALĒRE ‘be

9. There are examples in which it is the L-pattern allomorph which becomes dominant across the paradigm, e.g. *erguir* ‘to erect’ < ERIGO and *muñir* ‘to summon’ < MONEŌ.

worth', TRAHERE 'bring', CADERE 'fall' (cf. Spanish *hago, tengo, salgo, oigo, valgo, traigo, caigo*).^{10, 11, 12}

3. Velar allomorphy in Old Spanish: A purely morphological/morphomic account based on allomorphic endings

Morphemes are to be understood as a semantically heterogeneous collection of cells which, historically, were the locus of a large amount of allomorphy, usually on account of purely phonological factors. Once the phonological rules ceased to be a reality for the language, however, the allomorphy persisted in these cells and, over time, tended to converge upon a particular phonological exponence (Maiden, 2004). Thus, the different types of allomorphy in Table 7 converged upon velar allomorphy under the influence of Table 3–Table 6. Morphemes are created when a 'senseless' paradigmatic distribution of allomorphy becomes a grammatical reality for the language and a determinant of the distribution of the allomorphy. Thus, synchronically, the morpheme could be construed as a grammatical generalisation regarding the distribution of a common morphological formative, be it the lexical root or the desinential endings (see also the notion of a template, Aski, 1995).

10. The majority of scholars of Spanish (Menéndez Pidal, 1941, pp. 292–293; Penny, 2002, p. 174 amongst others) are of the opinion that the L-pattern reflexes of Latin FACIŌ 'do' derive from a form [fako] in which the *yod* was lost at a preliterary stage of the language. This view is due to the first ever attested Castilian L-pattern forms which display no effects of *yod*. See Malkiel (1974) for a refutation of this claim and a full discussion of the effects of *yod* on verbs in Ibero-Romance (see also O'Neill, 2012).

11. For TRAHERE and CADERE the presence of the *yod* must have been analogical since their Latin etyma did not contain a desinential *yod*. The insertion of a non-etymological *yod* must also be supposed in a number of other -ERE verbs whose root allomorph ended in /d/ (e.g. RĀDO 'scratch', CREDŌ 'believe', RŌDŌ 'gnaw', UADŌ 'wander'). For a discussion of the possible motivations of this analogical *yod* see O'Neill, 2012 and for an alternative view see Lloyd, 1987, pp. 296–297; Malkiel, 1974, p. 333; Penny, 2002, pp. 174–175. Likewise the reflexes of the Latin verb PETŌ 'aim at' and PONŌ 'put' in Portuguese (*peço, ponho*) both suggest a proto-form *[petjo] and *[ponjo].

12. Portuguese maintained a distinction between the results of /k/ + [j] and /k/ + front vowel, the first producing a voiceless sibilant the second a voiced one, FACIŌ > *faço*, DĪCIS > *dizes* (Williams, 1962, pp. 79 & 67). Spanish, however, underwent secondary voicing upon the results of /k/ + [j], thus ĒRĪCIU > *erizo*, DĪCIT > *dize* (Penny, 2002, p. 63 & 66). The reasons for this are uncertain and much debated (cf. Malkiel 1971, 1993 for a full discussion). Trusting Malkiel's conclusions, in what follows I shall assume that in preliterary Castilian /k/ + [j] > /tʃs/.

A purely morphological account of velar allomorphy in Spanish supposes that the cells of the L-pattern constitute a morpheme. Thus, the velar allomorphy is not licensed phonologically or semantically, but rather a purely morphological generalisation exists, that speakers are aware of, that a particular morphological formative, for example the root *caig-* of the verb *caer* 'fall' in Table 1, occurs in a specific set of cells. Such a conclusion is problematic since the cells of the L-pattern in Spanish can be reduced to a single phonological generalisation: the root of the verb is always followed by a non-front vowel. Such a correlation between an allomorph and a particular phonological environment invites a phonologically motivated interpretation of the allomorphy. Moreover, Bybee (1985) has argued that the distribution of the velar allomorph in Spanish could be determined by notions of markedness whereby the 1SG present indicative is considered to be the basic form from which all the other forms are derived (see Pérez-Saldanya, 1995 for similar arguments for Catalan).

Table 7. Reconstructed Ibero-Romance forms displaying etymological allomorphy in the L-pattern

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	[oujo]	[ouja]	[faʃo]*	[faʃa]	[βeɲo]	[βeɲa]	[trajo]	[traja]
2SG	[oues]	[oujas]	[faʃes]	[faʃas]	[βenes]	[βeɲas]	[traes]	[trajas]
3SG	[oue]	[ouja]	[faʃe]	[faʃa]	[βene]	[βeɲa]	[trae]	[traja]
1PL	[ouemos]	[oujamos]	[faʃemos]	[faʃamos]	[βenimos]	[βeɲamos]	[traemos]	[trajamos]
2PL	[ouetes]	[oujates]	[faʃetes]	[faʃates]	[βenites]	[βeɲates]	[traetes]	[trajates]
3PL	[ouen]	[oujan]	[faʃen]	[faʃan]	[βenen]	[βeɲan]	[traen]	[trajan]
	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	[teɲo]	[teɲa]	[sa.ʎo]	[sa.ʎa]	[βa.ʎo]	[βa.ʎa]	[kaɲo]	[kaɲa]
2SG	[tenes]	[teɲas]	[sa.ʎes]	[sa.ʎas]	[βa.ʎes]	[βa.ʎas]	[kaes]	[kaɲas]
3SG	[tene]	[teɲa]	[sa.ʎe]	[sa.ʎa]	[βa.ʎe]	[βa.ʎa]	[kae]	[kaɲa]
1PL	[tenemos]	[teɲamos]	[sa.ʎimos]	[sa.ʎamos]	[βa.ʎemos]	[βa.ʎamos]	[kaemos]	[kaɲamos]
2PL	[tenetes]	[teɲates]	[sa.ʎites]	[sa.ʎates]	[βa.ʎetes]	[βa.ʎates]	[kaetes]	[kaɲates]
3PL	[tenen]	[teɲan]	[sa.ʎen]	[sa.ʎan]	[βa.ʎen]	[βa.ʎan]	[kaen]	[kaɲan]

On the matter of the L-pattern forms being derived from the 1SG forms, O'Neill (2011a, pp. 204–246) has presented diachronic evidence from Spanish and other Romance varieties which contradict such a view (also see Wheeler, 2011 for Catalan). The conclusion of these authors is that there does not exist any coherent semantic set of features nor any morphological basic-derived relationship relating to frequency or markedness from which to derive the particular paradigmatic distribution of the velar allomorphs in Spanish.

On the topic of the velar allomorphy being derived phonologically, O'Neill (2011a, pp. 204–246) has advanced diachronic, synchronic, comparative and psycholinguistic evidence (Bybee & Pardo, 1981) which prompts the conclusion that the correlation between the velar allomorph and non-back vowels is a historical accident and in no way does the former determine the latter. As for the diachronic evidence, the author draws upon conclusions made by Martín Vegas (2007) which differ from other previous analyses of the spread of velar allomorphy in Spanish (Maiden, 1992, 2001; Malkiel, 1974; Menéndez Pidal, 1941, pp. 292–293; Penny, 2002, pp. 174–180) in claiming that, historically, the extension of the velar was an extension of the desinences *-go*, *-ga*, *-gas*, *-gamos*, *-gades*, *-gan*. For those verbs which possessed or adopted such desinences, it is impossible to claim that the velar consonant is phonologically licensed since, to take *poner* 'put', it is not the case that there are two allomorphs *pon-* and *pong-* whose distribution is determined phonologically; rather there is a lexeme PONER, with the root *pon*, which is marked as having the desinences *-go*, *-ga*, *-gas*, *-gamos*, *-gades*, *-gan* in the L-pattern cells.

The argument is the following: after the analogical changes whereby [fãʎo] > *fago* (see footnotes 11 and 12) there were a number of high-frequency verbs (*decir* 'say', *facer* 'do' and derivatives of DUCERE 'lead', see Table 3) which in the L-pattern cells ended in a velar consonant + /a/ + markers of person and number (*-go*, *-ga*, *-gas*, *-gamos*, *-gades*, *-gan*). These desinences coincided with the large number of inchoative verbs which, through regular sound change, displayed the same terminal elements, although in this case the velar consonant was voiceless (cf. *meres-co*, *flores-co*, *conos-co*, *nas-co*, *apares-co* of the type in Table 6).

Given such a situation, the suggestion is that the Old Castilian forms *nasco*, *aparesco*, *digo*, *fago*, *redugo*, *frango*, *cingo* were analysed as composed of a root *nas-*, *apares-*, *di-*, *fa-*, *redu-*, *fran-*, *cin-* plus the desinences *-go/-co*, *-ga/-ca*, *-gas/-cas*, *-gamos/-camos*, *-gades/-cades*, *-gan/-can* for the L-pattern cells. Martín Vegas (2007, p. 168) notes that

con un análisis de este tipo no hablaríamos de una inserción de *-g-* en determinados verbos, sino de la extensión de un esquema/modelo de flexión que caracteriza porque la 1ª persona del pres. ind. y el pres. subj. terminan en *-go*, *-ga*, *-gas* [...].

O'Neill (2011a, 2015) differs from this author in that he extends this analysis to motivate the velar element in some of the earliest attested velar verbs of Spanish *vengo*, *tengo*, *salgo* and relates it to the concept of the morpheme. With specific reference to the Castilian proto-forms [teɲo] < TENEŌ 'I have', [βeɲo] < VENIŌ 'I come', [saʎo] < SALIŌ 'I leave', if these were analysed as having the roots [teɲ-], [βeɲ-], [saʎ-], and were made to coincide with the morphological model above,

then such proto-forms would, through regular sound change, produce the attested words *tengo*, *vengo*, *salgo*, since the palatal consonants would be located in the coda of the syllable and would become depalatalised (cf. DOMINE > [doɲe] > *don* vs. DOMINA > *doña*, GALLU > *gallo* vs. GALLICU > *[gaʎgo] > *galgo* and also PIGNORA > *[peɲra] > *pendra* > *prenda*, Penny 2002, and COLLOCŌ > [kweʎgo] > *cuelgo*, Malkiel, 1974, p. 322).¹³

O'Neill (2011a, 2015) maintains that this analysis of the spread of velar allomorphy in Spanish is preferable to previous accounts due: (a) to its ability to explain the historical facts in a more convincing and less problematic way; and (b) to its ability to give an explanation to hitherto poorly justified changes in Spanish velar verbs.¹⁴ In what follows, these last changes will be briefly summarised since they not only support the hypothesis that in Old Spanish the velar element was parsed as part of the desinence but also offer a new perspective with which to explain the different diachronic developments of velar verbs within both Spanish and other varieties of Ibero-Romance.

13. Possible evidence for this analysis can be obtained from varieties of Aragonese which usually display a more extensive extension of the velar element. The phonological rules of these varieties differ regarding whether or not certain palatal consonants are allowed in the coda of the syllable or not, but in Ansotano (Barcos, 2007, p. 90) the palatal lateral [ʎ] is not allowed in coda positions (cf. BELLU > *bel*, BALLE > *bal*, Barcos, 2007, p. 37), and thus the verbs whose lexical root terminates in <ll> = [ʎ] display this throughout the paradigm with the exception of the lexemes which display the velar element (in *bullir* 'boil', *cullir* 'catch' and derivatives thereof) which in the L-pattern cells display the forms *bulg-*, *culg-*. In other varieties there is variation as to whether *cullir* displays the velar element (in Belsetán, cf. Lozano Sierra & Saludas Bernad, 2007, p. 86) and here the forms *cullo* alternate with *culgo*.

14. Firstly, this alternative analysis obviates the problems, first pointed out by Malkiel, inherent in explaining the origin of the velar forms in *tengo* and *vengo* via hesitation between the sounds /ɲ/ and /ŋg/ in verbs such as *tango* – *taño*. This alternation in this class of verbs, supposedly through hypercorrection, was paradoxically levelled in favour of the /ŋg/ sequence in those verbs in which the velar was *not* etymological and in favour of /ɲ/ in verbs in which it was (cf. Menéndez Pidal, 1941; Maiden, 1992, 2001). Secondly, this account can, in a straightforward way, explain why in verbs of the type *caigo*, *traigo* and *oigo*, the velar element is preceded by the glide [j], without the tenuous references of Malkiel (1974, pp. 335–336) to the double etymology of the forms of *traer* and the blending of the forms *trayo* and *trago* into *traigo* which acted as a catalyst for all other forms in *-ig-*. The explanation would be that these verbs were analysed as having the roots [kaj], [traj] and [oj] with which the endings [go], [gas], [ga], etc. were concatenated (cf. O'Neill, 2011a, 2015 for a full discussion).

4. Evidence for the velar element being morphologically parsed as belonging to the inflectional endings in Old Spanish but then possibly part of the root in Modern Spanish

The hypothesis that the extension of the velar consonant to other verbs in Spanish constituted an extension of desinences and not of velar allomorphs can provide an explanation for the irregular sound change that took place in the class of verbs which contained the Latin inceptive reflex -SK- in Spanish and possibly justify why in Portuguese, and some varieties of Asturian, these verbs lost their velar consonant after the medieval period. For Spanish, Penny (2002, p. 180) notes that, in the L-pattern cells, the Latin suffix -SK was realised as /sk/ (cf. MUSKA > *mosca* ‘fly’) due to it always being followed by a back vowel; in the non L-pattern cells, however, in which it was followed by a front vowel, the regular outcome was /f̥s/ (cf. PISCĒS > *peçes* > *peces* ‘fishes’); thus, Old Spanish present tense reflexes of the Latin verb COGNŌSCERE ‘know’, as shown in Table 8 along with the modern forms, display different allomorphs in the L-pattern than in the rest of the present tense (and the rest of the paradigm).

Table 8. Outcome of COGNŌSCERE*

	Old Spanish		Modern Spanish	
	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>conosco</i>	<i>conosca</i>	<i>conozco</i>	<i>conozca</i>
2SG	<i>conoçes</i>	<i>conoscas</i>	<i>conoces</i>	<i>conozcas</i>
3SG	<i>conoç(e)</i>	<i>conosca</i>	<i>conoce</i>	<i>conozca</i>
1PL	<i>conoçemos</i>	<i>conoscamos</i>	<i>conocemos</i>	<i>conozcamos</i>
2PL	<i>conoçedes</i>	<i>conoscades</i>	<i>conocéis</i>	<i>conozcáis</i>
3PL	<i>conoçen</i>	<i>conoscan</i>	<i>conocen</i>	<i>conozcan</i>

* Note that in Old Spanish the graph <ç> represented the affricate phoneme /f̥s/ and in Modern Spanish the graphs <z> and <c> are graphemes of the same sound /θ/ or /s/ depending on the variety.

Towards the end of the 15th century, however, the root final sibilant /s/ was replaced by the sound /f̥s/ (perhaps at this time modified to the dental sibilant /s̥/), which was characteristic of the majority of the forms of the paradigm (Penny, 2002, p. 108). This change cannot correspond to a process of levelling in favour of the lexical root of the majority of forms if the velar element is included within the root allomorph. If this were the case, the verbs would have two allomorphs: [konof̥s-] and [konosk-], and levelling would suppose the ousting of the latter, leaving behind a paradigm, well attested in the Asturian and Galician varieties, in

which the root is invariable [konoθ-]/[konos-] and there is no velar: thus standard Asturian, *conozo, conoces, conoce; conoza, conozas, conoza*, etc. I maintain that in Castilian the process of levelling did take place since, as opposed to some varieties of Asturian and Portuguese, the root allomorphs were not [konot̪s-] and [konosk-] rather [konot̪s-] and [konos-], and the velar element was included as part of the ending. Accordingly, the effect of levelling produces the attested Modern Castilian forms in Table 8 in which the graphs <z> and <c> are graphemes of the same sound /θ/.

Another advantage of considering the velar consonant as being parsed as part of the desinence in Old Spanish is that it can account for the loss of the velar consonant in lexemes in which it was etymological (TANGŌ > *tango* >> *taño* 'I play', FRANGŌ > *frango* >> *fraño* 'I break', CINGŌ > *cingo* >> *ciño*, TINGŌ > *tengo* >> *tiño* 'I dye', SPARGŌ > *espargo* >> *esparzo* 'I scatter', RINGŌ(R) > *ringo* >> *riño* 'I scold'). My explanation for the loss of the etymological velar is that these verbs did not and could not correspond to the general morphological structure for velar L-pattern forms.

There is ample data from the history of the Romance languages (Maiden, 2001, 2004, 2009) which testify to the tendency of allomorphy in morphemes to converge upon a particular phonological exponence. I contend that verb forms can also converge on a particular morphological structure. For L-pattern velar verbs in Old Spanish I claim that this structure was the root of the rhizotonic forms of the present indicative plus the velar endings (-go, -ga, -gas, etc.). There were a number of verbs which displayed this structure, especially those verbs in Table 6 which contained the Latin ingressive suffix. These verbs had a high type frequency and constituted a productive way in which to form adjectival and nominal verbal derivations, often via parasynthesis (*enorgullecer* 'make proud', *ensordecer* 'make deaf'). This morphological structure also coincided with a number of lexemes which contained a newly acquired voiced velar consonant, and which had a high token frequency: *cae* – *caiga*, *trae* – *traiga*, *o(y)e* – *oiga*, *sale* – *salga*, *vale* – *valga*, *pone* – *ponga*. There were, of course, exceptions to this rule, notably *dice* – *diga*, *tiene* – *tenga*, *viene* – *venga* but, despite the token frequency of these verbs, such alternations were not found elsewhere. Therefore, upon the assumptions that: (a) the most prominent morphological structure for velar verbs in the L-pattern was for them to consist of a rhizotonic present indicative root to which velar desinences were concatenated; and (b) that morphemes can converge upon a certain type of transparent morphological structure, just as they converge upon a phonological exponent; then the levelling of allomorphy in verbs of the type in Table 5, e.g., *cingo* > *ciño*, *tango* > *taño*, *plango* > *plaño*, is relatively straightforward.

These verbs posed problems for the L-pattern since even though their endings did conform to the tendency towards velar allomorphy, their morphological

structure did not conform to the pattern detailed above since the forms *cin-go*, *fran-go*, *tan-go*, *plan-go* did not correspond to the preferred structure whereby the stem of the L-pattern was based on the root of the other rhizotonic forms of the present indicative (*ciñ-*, *frañ-*, *tañ-*, *plañ-*) plus the velar endings. Moreover, such forms could not come about via levelling of the root allomorph since the desired forms **ciñgo*, **frañgo*, **tañgo*, **plañgo* were phonologically impossible given the restriction against palatal consonants in coda position (DOMINE > [dope] > *don* vs. DOMINA > *doña*, GALLU > *gallo* vs. GALLICU > *[gaʎgo] > *galgo* and also PIGNORA > *[pepra] > *pendra* > *prenda* (Penny, 2002) and COLLOCŌ > [kweʎgo] > *cuelgo* (Malkiel, 1974, p. 322). The verbs of the type in Table 3, however, which also displayed a ‘non-preferred’ structure in the L-pattern cells, i.e. the root upon which the velar endings were concatenated was not the same as that of the rest of the present indicative (*redu-go* vs. *reduc(e)*, *adu-go* vs. *aduc(e)*) could readily undergo levelling since they did not contain a final palatal consonant: *redugo* > *reduzgo* > *reduzco*; *adugo* > *aduzgo* > *aduzco*.¹⁵ Since such an option was not available to the verbs with root final palatal nasal consonants the optimal option was for these verbs not to be included in the group of verbs with velar endings thus *cin-go* > *ciñ-o*, *tan-go* > *tañ-o* (see also Pérez-Saldanya, 1995, p. 420).

However, in its origins, the velar element must have been considered part of the lexical root in Spanish, hence the development whereby FACIŌ > *fago* and IACIŌ > *iago* and not the expected [faʃsko] and [jaʃsko], based on the assumption that the velar element consisted of a concatenation of velar endings. The attested forms are a matter of proportional analogy with the roots of present tense forms of verbs of the type [dik-o] – [diðz-e]. Such verbs, which had an alternation between a root allomorph ending in /g/ in the L-pattern and [dʒ] elsewhere, had a high type and token frequency: the verb *decir* ‘say’, reflexes of verbs derived from DŪCERE (see Table 3 REDŪCŌ ‘I reduce’ and ADDŪCŌ ‘I adduce’) and also verbs of the type in Table 4 (SPARGŌ > *espargo* ‘I scatter’ and ERIGŌ > *yergo* ‘I erect’).¹⁶ Moreover, it is far from clear, for Modern Spanish, that velar allomorphy is a function of the endings of verbs and not their roots. This

15. It must be noted, however, that such a development is not attested for *espargo*.

16. Moreover, this hypothesis gains strength when one takes into account the relative productivity that the alternation [g] – [dʒ] must have had in the Late Latin/Early Romance of the Iberian Peninsula on account of the number of semantically related derivatives [noðze] ‘nut’ – [nokale] ‘nut tree’ (see Martín Vegas, 2007, pp. 117–173). Much of these derivatives, however, in Modern Spanish have become semantically opaque, thus: *paz* ‘peace’, *pagar* ‘to pay’, *hoz* ‘sickle’ and *ahogar* ‘to drown’ (Martín Vegas, 2007, pp. 163–166).

uncertainty is prompted by the fact that, with the exception of verbs of the type *merecer* 'be worth', which display the voiceless velar consonant, velar allomorphy in Spanish has, from the Golden Age onwards, become reduced in number and is now limited to a small number of verbs which generally display other allomorphy within their paradigms.¹⁷ There is no evidence that the velar allomorphy of the type *salgo, valgo, tengo, vengo, traigo* for *salir* 'go out', *valer* 'be worth', *tener* 'have', *venir* 'come' and *traer* 'bring' is the result of the concatenation of a particular type of velar allomorphic ending, as opposed to another root allomorph for these forms. That is, it is not clear that the morphological segmentation of *salgo* is *sal-go* and not *salg-o*, especially when one considers the amount of root allomorphy which such forms display elsewhere (cf. the allomorphs *saldré, valdré, tiene, viene, trajera*).

I therefore assume the hypothesis that there can be variation and change as to the morphological parsing of the velar element. In what follows, I shall provide evidence in support of this hypothesis; the majority of the evidence will involve how velar allomorphy interacts with diphthongised roots in central Ibero-Romance varieties. Firstly, however, it is necessary to introduce another morpheme of Ibero-Romance, the N-pattern morpheme, since the allomorphy determined by this morpheme, that of diphthongised stems, overlaps with the L-pattern morpheme and the scope of each morpheme (i.e. whether they can be conceived as generalisations regarding the endings or roots/stems of verbs) is revealing as to the morphological parsing of the velar element.

5. The N-pattern morpheme

The label 'N-pattern' refers to a pattern of alternation, recurrent across the Romance languages, whereby the forms of the first, second and third persons singular and third person plural of the present indicative and of the present subjunctive, and the second person singular of the imperative, share a distinctive common form, usually a common root allomorph.¹⁸ This is illustrated in Table 9 with examples from the Italian verbs *morire* 'die', *udire* 'hear' and *dovere* 'ought to'.

17. Other verbs of type of *merecer* are *vencer, cocer, escocer, ejercer, torcer*.

18. In subsequent publications Maiden (2011) has stated that it is not merely a common allomorph but rather common characteristics of various types which are not found elsewhere in the paradigm.

Table 9. The N-pattern morphome for *morire*, *udire* and *dovere*

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>muoio</i>	<i>muoia</i>	<i>odo</i>	<i>oda</i>	<i>devo</i>	<i>deva</i>
2SG	<i>muori</i>	<i>muoia</i>	<i>odi</i>	<i>oda</i>	<i>devi</i>	<i>devas</i>
3SG	<i>muore</i>	<i>muoia</i>	<i>ode</i>	<i>oda</i>	<i>deve</i>	<i>deva</i>
1PL	<i>moriemo</i>	<i>moriemo</i>	<i>udiamo</i>	<i>udiamo</i>	<i>dobbiamo</i>	<i>dobbiamo</i>
2PL	<i>morite</i>	<i>moriato</i>	<i>udite</i>	<i>udiate</i>	<i>dovete</i>	<i>dobbiate</i>
3PL	<i>muoiono</i>	<i>muoiano</i>	<i>odono</i>	<i>odano</i>	<i>devono</i>	<i>devano</i>
2SG	Imper.	<i>muori</i>	Imper.	<i>odi</i>	Imper.	–
1SG	Imp. Indic.	<i>morivo</i>	Imp. Indic.	<i>udivo</i>	Imp. Indic.	<i>dovevo</i>

The cells of the N-pattern also share the property of being rhizotonic. Maiden (2009, 2011), however, considers the allomorphy in these stems to be neither phonologically derivable nor conditioned (however cf. Anderson 2008, 2010 and 2011 for an alternative view for the Rumantsch variety of Surmiran, and Maiden, 2011 for a response). Such a conclusion is problematic for Spanish since, in Modern Spanish, the only type of unambiguous N-pattern allomorphy is that of diphthongised roots, as shown in Table 10 for *perder* ‘lose’ and *mover* ‘move’, and this allomorphy is largely considered to be a matter of phonological conditioning from an underlying invariant form (Harris, 1969, 1977, 1978, 1985; Schuldberg, 1984; García-Bellido, 1986; Carreira, 1991).

Table 10. The N-pattern for *perder* and *mover*

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>pierdo</i>	<i>pierda</i>	<i>muevo</i>	<i>mueva</i>
2SG	<i>pierdes</i>	<i>pierdas</i>	<i>mueves</i>	<i>muevas</i>
3SG	<i>pierde</i>	<i>pierda</i>	<i>mueve</i>	<i>mueva</i>
1PL	<i>perdemos</i>	<i>perdamos</i>	<i>movemos</i>	<i>movamos</i>
2PL	<i>perdáis</i>	<i>perdáis</i>	<i>movéis</i>	<i>mováis</i>
3PL	<i>pierden</i>	<i>pierdan</i>	<i>mueven</i>	<i>muevan</i>

O’Neill (2011a, pp. 246–279), however, has advanced diachronic, synchronic, psycholinguistic and cross-linguistic evidence to suggest that the diphthong-monophthong alternation which exists in a great many verbs in Spanish and which is correlated with word stress is not a matter of an invariant underlying form from which the different allomorphs are predictable via a phonological stress conditioned rule (see also Bermúdez-Otero, ms.). Rather, the different allomorphs must

be considered to be lexically stored. Furthermore, he has argued (contra assumptions of Bermúdez-Otero, ms.) that the selection of the respective allomorphs is not a matter of phonologically conditioned allomorphy whereby rhizotonic stress selects for the diphthongised allomorph, since word stress itself in the Spanish verb cannot be predicted on the basis of phonological criteria (*contra* Roca, 1990). The most straightforward solution is to admit that the different stored allomorphs are merely lexically marked as being stressed or else that the allomorphs are marked to occur in the N-pattern and that stress in the Spanish verb is morphologically determined.

The N-pattern therefore is a morpheme for Spanish and I would claim that this also potentially holds for all Ibero-Romance varieties. As to whether the morpheme is to be considered a generalisation over roots or desinences, there is comparative evidence to suggest that both options are possible. Thus, the verbs of what is traditionally termed the Catalan IIIa conjugation display the increment *-eix*, exclusively in the N-pattern cells as illustrated in Table 11 for the Catalan verb *servir* 'serve'.¹⁹

Table 11. The N-pattern for *servir*

	Pres. Indic.	Pres. Subj.
1SG	<i>serveixo</i>	<i>serveixi</i>
2SG	<i>serveixes</i>	<i>serveixis</i>
3SG	<i>serveix</i>	<i>serveixi</i>
1PL	<i>servim</i>	<i>servim</i>
2PL	<i>serviu</i>	<i>serviu</i>
3PL	<i>serveixen</i>	<i>serveixin</i>

In Modern Catalan, this increment, as its name suggests, is generally considered to be a regular formative which is concatenated to the verbal root and which constitutes a regular model of verbal inflection for a group of verbs (Perea, 2002, p. 600).²⁰ Thus, the morphological parsing is *serv* + *eix*. There are other verbs,

19. Etymologically the formative *-eix* derives from the Latin inchoative or ingressive affix -SC- e.g. FLÖRET 'it flowers' vs. FLÖRĒSCET 'it's coming into bloom'. In Latin, this augment was incompatible with perfective meaning and thus could only appear in imperfective verb forms. In a number of Romance varieties, however, what had then become an unmotivated distribution of allomorphy was incorporated into an N-pattern distribution.

20. Mascaró (1983) has even suggested that it is an allomorph of the thematic vowel.

however, for which the N-pattern distribution implies a different selection of a lexical allomorph; witness the present tense paradigm for the verb *anar* ‘go’ in Table 12 (cf. the present indicative form of the verb *ir* in Old Spanish: *vo, vas, va, imos, ides, van*).²¹

Table 12. The N-pattern for *anar*

	Pres. Indic.	Pres. Subj.
1SG	<i>vaig</i>	<i>vagi</i>
2SG	<i>vas</i>	<i>vagis</i>
3SG	<i>va</i>	<i>vagi</i>
1PL	<i>anem</i>	<i>anem</i>
2PL	<i>aneu</i>	<i>aneu</i>
3PL	<i>van</i>	<i>vagin</i>

The foregoing suggests that there can be variation as to what morphological formatives morphemes are sensitive to with regards to formal identity; that is, morphemes could be considered to be generalisations regarding the distribution of a certain root allomorph or of a particular desinential formative. This conclusion becomes interesting when one takes into account that there can be variation as to the morphological parsing of the velar element and that the N-pattern and L-pattern morphemes overlap. The L-pattern morpheme involves the 1SG present indicative and all the present subjunctive cells, whilst the N-pattern involves the 1SG, 2SG, 3SG and 3PL present indicative and subjunctive as well as the 2SG imperative. This is illustrated in Table 13 in which the medium-tone grey shading denotes cells which are exclusively part of the N-pattern, the dark grey shading marks cells which are exclusively part of the L-pattern, the light grey designate those cells which are common to both patterns and the clear cells mark those forms which do not form part of either morpheme.

21. Historically, these verbs are cases of inceptive suppletion (used in the sense of Corbett, 2007). Inceptive suppletion is where two or more different lexemes are integrated into a paradigm of a single lexeme, usually in mutually exclusive cells. Thus, Catalan contains reflexes of Latin UĀDERE ‘wander’, surviving exclusively in the N-pattern cells but reflexes of AMBULĀRE ‘walk’ exclusively in the remainder of the paradigm. In Old Spanish the reflexes were of the paradigms of UĀDERE and ĪRE ‘to go’.

Table 13. The L-pattern and N-pattern overlap

	Pres. Indic.	Pres. Subj.
1SG		
2SG		
3SG		
1PL		
2PL		
3PL		
	Imperative	
2SG		
2PL		

Given the assumption that morphemes can determine the distribution of a morphological formative (be it a root, infix or ending), this overlap of morphemes poses problems for cases in which diphthongised roots co-occur with velar allomorphs since usually the occurrence of a particular formative (e.g., diphthongised stem or velar element) in one cell of the morpheme implies its presence in all the cells of the morpheme (Maiden, 2004).²² Thus, the cells marked with light grey shading are 'torn' between respecting the formal coherence of identical allomorphy of the N-pattern or the L-pattern. This matter is further complicated depending on the morphological segmentation of the velar element.

In the following sections, I will provide evidence from different varieties of central Ibero-Romance which show a number of solutions to this clash of morphemes. The data supports the hypothesis that there can be variation as to how the velar element can be morphologically parsed and suggests that morphemes can not only converge upon one particular type of phonological form (Maiden, 2004) but also on a particular morphological structure or type.

6. The clash of the morphemes

The first hypothetical solution to the clash of the morphemes is that when morphemes overlap, the tendencies of coherence of each particular morpheme can be focused upon a particular type of phonological and/or morphological exponent.

22. This is however a tendency and diagnostic of morphemes rather than a defining principle since it is not always the case that a particular allomorph needs to be present in all cells of the morpheme (see *se* vs. *sepa* from Spanish *saber* 'know'). See also arguments in Wheeler (2011).

Each morpheme is sensitive to a different morphological formative: the N-pattern to the lexical root characterised by diphthongs and the L-pattern to velar endings. Such is the situation that I envisage may have held in Old Spanish since diphthongised roots were compatible with velar allomorphy and, even though there was variation as to which lexemes diphthongised and which presented velar allomorphy, the distribution of the former was the N-pattern and the latter the L-pattern cells (see *tuele*, *tuelga*, *tolgamos*, *tolgades* for the verb *toler* ‘take’). Such a state of affairs also holds, in my opinion, in the Asturian variety of Cándamo (Díaz González, 1986, p. 81) as attested by the reflex of DOLÈRE > *doler* – *duler* in Table 14 and also in the Ansotano variety of Aragonese (Barcos, 2007, p. 90), as illustrated by the reflexes of TORQUÈRE > *torzer* and TENÈRE > *tener* in Table 15.

Table 14. Asturian (Cándamo) outcome of DOLÈRE

	Pres. Indic.	Pres. Subj.
1SG	<i>duelgo</i>	<i>duelga</i>
2SG	<i>duelis</i>	<i>duelgas</i>
3SG	<i>duel</i>	<i>duelga</i>
1PL	<i>dulemos</i>	<i>dulgamos</i>
2PL	<i>duleis</i>	<i>dulgais</i>
3PL	<i>duelin</i>	<i>duelgan</i>

Table 15. Aragonese (Ansotano) outcome of TORQUÈRE and TENÈRE

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>tuerzco</i>	<i>tuerzcai</i>	<i>tiengo</i>	<i>tiengai</i>
2SG	<i>tuerces</i>	<i>tuerzcas</i>	<i>tienes</i>	<i>tiengas</i>
3SG	<i>tuerce</i>	<i>tuerzca</i>	<i>tiene</i>	<i>tienga</i>
1PL	<i>torcemos</i>	<i>torzcamos</i>	<i>tenemos</i>	<i>tengamos</i>
2PL	<i>torcez</i>	<i>torzaz</i>	<i>tenez</i>	<i>tengaz</i>
3PL	<i>tuercen</i>	<i>tuerzcan</i>	<i>tienen</i>	<i>tiengan</i>

These data are compatible with the assumption that the N-pattern is the domain of diphthongisation and of morphological roots and the L-pattern that of velar allomorphy and desinential endings.²³

23. Any phonologist, however, could claim that given the fact that diphthongisation is coterminous with stress and velar allomorphy with the appearance of a non-back vowel, then the allomorphy is phonologically licensed. This, of course, is a possibility, but, as demonstrated for

Independent proof in support of such an hypothesis can be found in other velar verbs from the Ansotano variety which do not display diphthongised roots but for which the L-pattern forms seem to be constructed on the N-pattern roots (*traí – traigo, fui – fuigo, bei – beigo*). Further evidence for such a conclusion, not involving diphthongising verbs, comes from the Aragonese locality spoken around the valley of Benasque (see Saura Rami, 2003, pp. 233–271). The speech of this area is considered to be transitional between Aragonese and Catalan varieties and the different localities exhibit a great deal of morphological variation. With respect to velar allomorphy, the distribution of this allomorphy is, as with other varieties of Aragonese and as opposed to Catalan, exclusively in the L-pattern cells. Of particular interest to the present discussion is that the L-pattern forms seem to be built via the concatenation of a velar consonant with the root of the 3sg present indicative. This is illustrated in Table 16 for the verbs *veure* ‘live’, *creure* ‘believe’ and *beure* ‘drink’. Note that the paradigms given here do not correspond to those given by the author (*ibid.*, pp. 233–271) since the author tries to incorporate all the variation in his paradigms. The verb forms below correspond to those from Zone B (see Saura Rami, 2003, p. 28 for a list of these localities) based on the supplementary information regarding the root allomorphs and desinences in this zone (*ibid.*, pp. 215–222).

Spanish (O'Neill, 2011a, pp. 211–246), the mere coincidence of an allomorph and a phonological environment does not necessarily imply that there exists a causal relationship between both. The general assumption in modern phonological theory is, however, that such recurrent phonological correlations do imply a causal relationship; Burzio (2004, p. 38) on the topic of velar allomorphy in Italian states that “whatever identity relations have a statistical presence in the data, also have, *ipso facto*, a grammatical status, expressible as a faithfulness constraint in the O[ptimality] T[heory] formalism”. I question this assumption and follow Eddington (2004, p. 3) who has pointed out that “a detailed, rigorous, or sophisticated description of a linguistic phenomenon does not necessarily indicate that the phenomenon has any relevance to linguistic cognition” (see also Goyvaerts, 1978; Lass, 1976; Morin, 1988; Skousen, 1989). In fact, the assumption that a recurrent phonological correlation between a phonological and morphological element is tantamount to phonologically conditioned allomorphy constitutes the promotion of linguistic description to the status of psychological explanation (Black & Chiat, 1981, p. 48) and in doing so, creates a circularity in the argumentation since the data/observable facts which need to be explained are equated with the very explanation of the data (see Eddington, 2004, p. 14; Higginbotham, 1991, p. 555; Ohala, 1990, p. 159; Sampson, 2001, p. 124). Given the lack of any independent evidence to suggest that the allomorphy discussed in this section is phonologically conditioned and on the basis of the comparative Spanish data I entertain the hypothesis that the allomorphy is morphologically conditioned. From such a perspective the evidence from the paradigms above indicate that the root of the L-pattern is created by the concatenation of a velar ending to the root of the N-pattern.

Table 16. The present tense for *veure*, *creure* and *beure* in Benasque (Aragón)

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>viugo</i> ~ <i>vivo</i>	<i>viuga</i> ~ <i>viva</i>	<i>creugo</i> ~ <i>credo</i>	<i>creuga</i> ~ <i>creda</i>	<i>beugo</i> ~ <i>bebo</i>	<i>beuga</i> ~ <i>beba</i>
2SG	<i>vius</i>	<i>viuguas</i> ~ <i>vivas</i>	<i>creus</i> *	<i>creguas</i> ~ <i>credas</i>	<i>beus</i>	<i>beguas</i> ~ <i>bebas</i>
3SG	<i>viu</i>	<i>viuga</i> ~ <i>viva</i>	<i>creu</i>	<i>cregua</i> ~ <i>creda</i>	<i>beu</i>	<i>begua</i> ~ <i>beba</i>
1PL	<i>vivim</i>	<i>viuga</i> ~ <i>vivam</i>	<i>credem</i>	<i>creguan</i> ~ <i>credam</i>	<i>bebem</i>	<i>beugam</i> ~ <i>bebam</i>
2PL	<i>vivits</i>	<i>viuguats</i> ~ <i>vivats</i>	<i>credets</i>	<i>creguats</i> ~ <i>credats</i>	<i>bebets</i>	<i>beuguats</i> ~ <i>bebats</i>
3PL	<i>viven</i>	<i>viuguan</i> ~ <i>vivan</i>	<i>creden</i>	<i>creguan</i> ~ <i>credan</i>	<i>beben</i>	<i>beuguan</i> ~ <i>beban</i>

* For the 2SG and 3SG present indicative the variants *credes*, *crede* respectively are given in parenthesis in the original. Presumably this notation means that these forms are attested but very infrequent.

Here, in the L-pattern, the etymological roots *viv-*, *cred-*, and *beb-* alternate with the velar forms *viug-*, *creug-*, *beug-*.²⁴ The forms *viu*, *creu* and *beu* are etymologically expected and attested in 3SG forms of the present indicative via the vocalisation of a final labial, dental or voiceless velar consonant after apocope had taken place (*esriu* < *scribe* < SCRIBIT, cf. *dèu* < DECER, *pèu* < PEDE, *nèu* < *NEVE, *perdiu* < PERDICE, *cllau* < CLAVE, *ibid*, p. 105). This sound change, although not explicitly stated, must have also held in coda final position (SCRIBERE > *scribe* > *escriure*). Apocope did not always occur with final /o/ (cf. *güello*, *sapo*, *bllanco* and derivations in *-ano*, *-iello* and *-uelo* (*ibid*, p. 64)²⁵ and thus, the alternative morphological forms for the 1SG present indicative have come about either by way of the concatenation of the endings *-go*, *-ga*, *-gas*, *-gam*, *-gats*, *-gan* to the form of the 3SG *viu* + *go* > *viugo*, or, assuming the validity of the vocalisation of the consonants in coda position, via the concatenation of the aforementioned endings to the root allomorph of the 1SG (*vid-*, *cred-* and *beb-*) producing the forms *vivgo* > *viugo*, *credgo* > *creugo*, *bebgo* > *beugo*.

Upon the assumption that the allomorphy is not phonologically determined (see footnote 23), the varieties analysed in this section (those of Table 14 and Table 15) suggest that the velar element forms part of the desinential ending and the consistency of the velar consonant is determined by the L-pattern, whilst the

24. This is also the case with the verbs *riure* 'laugh', *chaure* 'lie', *deure* 'owe', *escriure* 'write' in which the 1SG present indicative forms have the respective alternants: *rido* – *riugo*, *chaso* – *chaugo* – *chasco*, *debo* – *deugo*, *escribo* – *escriugo*.

25. Note however that in the plural form the /o/ is lost: *sapo* – *saps*, *bllanco* – *bllancs*, *casuelo* – *casuels*. Also, final /o/ is sporadically lost under the influence of neighbouring Catalan varieties (*llop* < LUPU, *caball* < CABALLU, *llech* < LECTU, *ibid*, p. 63).

distribution of the diphthongised root allomorphs is determined by the N-pattern. Such evidence not only shows that morphemes can be sensitive to different morphological formatives but it also confirms their tendency towards convergence on a particular phonological *and* morphological form.

As stated, I assume such a state of affairs held for Old Spanish as attested by lexemes in which velar allomorphy and diphthongisation coincided either due to regular sound change (*cuego/cuezo* 'I cook') or analogy (*duelgo* 'I hurt', *muelgo* 'I grind', *tuelgo* 'I take away', *suelgo* 'I am accustomed to', *fiergo* 'I harm'); Modern Spanish has, in all cases, preferred the diphthongised forms and eradicated the velar allomorphs. In fact, it is a peculiarity of Castilian, in comparison with its central Ibero-Romance neighbours, that there are no verbs in which diphthongisation coincides with velar allomorphy. This, I claim, is because in Spanish velar allomorphy is no longer exclusively a function of the endings of verbs but could be viewed as an allomorph of the root, at least for verbs with voiced velar allomorphy. That is, the observed mutually exclusive relationship between velar allomorphy and diphthongs suggests that speakers analysed them as alternative manifestations of root allomorphy and not one type of allomorphy pertaining to roots and the other to endings. In such a case, there occurred a clash of morphemes and a competition over forms: the velar root *duelg-* of the verb *doler* was present in all the L-pattern cells but the root *duel-* was only present in a subsection of the N-pattern cells. Given the type frequency of diphthongising verbs, together with the inclusion of the most frequent and least marked form of the paradigm, 3SG present indicative, in the N-pattern, it is not surprising that this pattern won out in Spanish and thus *duelga* > *duela*. In neighbouring Romance varieties, especially those to the East, which contained more lexemes displaying velar allomorphy, this solution was not dominant and what could occur is that both patterns could exist but one was able to take precedence over the other. The L-pattern could dominate the N-pattern (L>>N, to be read as the L-dominates-N-pattern) or vice versa, the N-pattern could dominate the L-pattern. In the following section I will only analyse the cases in which the L-pattern dominates the N-pattern (L>>N pattern) since the converse is not attested for velar allomorphy, although it is arguably attested for Spanish for vocalic allomorphy (O'Neill 2011a, pp. 339–349, 2011b).²⁶

26. Such a pattern would present a diphthongised roots in the N-pattern (*duelo*, *duele*, *duela*) and the velar allomorphy would be relegated only to the 1PL and 2PL L>>N (*dolgamos*, *dolgáis*) present subjunctive, the only cells of the L-pattern which do not overlap with the N-pattern.

6.1 The creation of the L>>N-pattern

In the previous section I presented data from Asturian and Aragonese varieties for which I argued that: (a) velar allomorphy was applied to endings and was distributed according to the L-pattern; and (b) that diphthongised stems were a question of allomorphic roots and distributed according to the N-pattern (I also defended this for Old Spanish). Such a state of affairs does not hold, however, in the Aragonese varieties of the Alta Ribagorza (Haensch, 2003, pp. 141–143), since when diphthongisation and velar allomorphy co-occur, the velar is, I maintain, analysed as part of the lexemic root together with the diphthong and thus, in accordance with the tendency towards convergence in the L-pattern, the diphthong spreads to all the other cells of the L-pattern in which it is not present: the arhizotonic 1PL and 2PL present subjunctive forms. This is illustrated by the verbs *tórse* ‘twist’ and *tínrrre* ‘have’ in Table 17 (Haensch, 2003, pp. 128, 121–122).

Table 17. The L>>N-pattern for *tórse* and *tínrrre*

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>twérsko</i>	<i>twérska</i>	<i>tjéngo</i>	<i>tjénga</i>
2SG	<i>twérses</i>	<i>twérskas</i>	<i>tjénes</i>	<i>tjéngas</i>
3SG	<i>twérese</i>	<i>twérska</i>	<i>tjéne</i>	<i>tjénga</i>
1PL	<i>torsém</i>	<i>twerskám</i>	<i>tením</i>	<i>tjengám</i>
2PL	<i>torséts</i>	<i>twerskátz</i>	<i>teníts</i>	<i>tjengáz</i>
3PL	<i>twérsen</i>	<i>twérskan</i>	<i>tjénen</i>	<i>tjéngan</i>

In this variety, the L-pattern dominates the N-pattern, and the latter is reduced to the 2SG, 3SG and 3PL present indicative (the L>>N-pattern). Such domination of the N-pattern by the L-pattern is only valid for those lexemes in which velar allomorphy and diphthongisation coincide. Lexemes which only display diphthongised allomorphs distribute this allomorphy, according to the N-pattern only, as illustrated by *poder* ‘be able’ (Haensch, 2003, p. 130) and *dormir* ‘sleep’ (ibid, p. 145) in Table 18.

In this variety, therefore, the N-pattern still retains its original, and in many cases etymological, distribution associated with the class of lexemes which only display diphthongised allomorphs.²⁷ However, when this type of allomorphy

27. Such a situation also holds in the Cantabrian Valle del Pas (Penny, 1970), although in this variety there is variation as to whether the diphthong appears in the 1PL and 2PL present subjunctive with velar verbs only (*cozcámos* ~ *cuezcámos*, *cozcáis* ~ *cuezcáis*, ibid, p. 126).

Table 18. The N-pattern for *poder*

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>pwédo</i>	<i>pwéda</i>	<i>dwérmo</i>	<i>dwérma</i>
2SG	<i>pwéts</i>	<i>pwédas</i>	<i>dwérmes</i>	<i>dwérmas</i>
3SG	<i>pwéde</i>	<i>pwéda</i>	<i>dwérme</i>	<i>dwérma</i>
1PL	<i>podém</i>	<i>podám</i>	<i>dormim</i>	<i>dormám</i>
2PL	<i>podéts</i>	<i>podátz</i>	<i>dormíts</i>	<i>dormátz</i>
3PL	<i>pwéden</i>	<i>pwédan</i>	<i>dwermen</i>	<i>dwérman</i>

coincides with velar allomorphy there occurs a conflict between the coherent tendency for there to appear the same allomorph over both the N-pattern cells and L-pattern cells. These data demonstrate that different morphemes can be correlated with different classes of lexemes depending on the type of allomorphy which they exhibit. Thus, in the variety under scrutiny, velar allomorphs are distributed according to the L-pattern, diphthongised allomorphs according to the N-pattern but when both coincide the result is an L>>N-pattern.

7. Conclusions and implications

In the foregoing I have, on the basis of previous studies (O'Neill, 2011a, 2015; Wheeler, 2011), assumed that the distribution of velar allomorphy and diphthongisation in Ibero-Romance is morphologically determined by reference to what have been termed morphemes (Aronoff, 1994; Maiden, 2004). I have presented data from verbs displaying such allomorphy from different varieties of Ibero-Romance and at different diachronic stages. The comparative results prompt the conclusion that: (a) there can be variation and change as to how the velar element is morphologically parsed; and (b) what morphological formatives morphemes can be construed as being generalisations about (roots or desinences or formatives which may be classed as neither). The generalisations which emerge from the data are that there is a tendency not only for morphemes to be correlated with a particular phonological type of allomorphy (velar allomorphy, diphthongised stems or a combination of both) but also with a particular morphological type (root allomorphy, desinential allomorphy) or structure which tends to become transparent and associated with a particular abstract paradigmatic patterning and also a particular phonological form. Thus, in Old Castilian the N-pattern was a generalisation about diphthongised roots (*duel(e)*, *cuez(e)*), the L-pattern about velar endings and both were compatible (*tuelga*, *duelga*, *cuezga*), producing a tendency towards a structure whereby the root of the N-pattern formed the basis for the

inflectional velar forms of the L-pattern via the concatenation with velar endings. The tendency towards this structure can explain the loss of the etymological velar verbs of the type *cingo*, *frango* and the hybrid Golden Age forms *reduzgo* < *redugo*, *aduzgo* < *adugo* (see Section 4). Moreover, the loss of the velar element in Latin ingressive verbs in some varieties of Asturian and Portuguese versus its maintenance in Castilian could be explained via the different parsing of the velar element; in Portuguese it was considered as part of the root allomorph.

The L-pattern in Portuguese stands out amongst the Ibero-Romance languages in that it allows various types of allomorphy and has not undergone convergence upon one particular phonological type: witness the modern forms of the Portuguese verbs *fazer* ‘do’, *medir* ‘measure’, *valer* ‘be worth’, *vir* ‘come’, *ver* ‘see’, *caber* ‘fit’ in Table 19.

Table 19. The L-pattern in Portuguese

	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>faço</i>	<i>faça</i>	<i>meço</i>	<i>meça</i>	<i>valho</i>	<i>valha</i>
2SG	<i>fazes</i>	<i>faças</i>	<i>medes</i>	<i>meças</i>	<i>vales</i>	<i>valhas</i>
3SG	<i>faz</i>	<i>faça</i>	<i>mede</i>	<i>meça</i>	<i>vale</i>	<i>valha</i>
1PL	<i>fazemos</i>	<i>façamos</i>	<i>medimos</i>	<i>meçamos</i>	<i>valemos</i>	<i>valhamos</i>
2PL	<i>fazeis</i>	<i>façais</i>	<i>medis</i>	<i>meçais</i>	<i>valeis</i>	<i>valhais</i>
3PL	<i>fazem</i>	<i>façam</i>	<i>medem</i>	<i>meçam</i>	<i>valem</i>	<i>valham</i>
	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.	Pres. Indic.	Pres. Subj.
1SG	<i>venho</i>	<i>venha</i>	<i>vejo</i>	<i>veja</i>	<i>caibo</i>	<i>caiba</i>
2SG	<i>vens</i>	<i>venhas</i>	<i>vês</i>	<i>vejas</i>	<i>cabes</i>	<i>caibas</i>
3SG	<i>vem</i>	<i>venha</i>	<i>vê</i>	<i>veja</i>	<i>cabe</i>	<i>caiba</i>
1PL	<i>vimos</i>	<i>venhamos</i>	<i>vemos</i>	<i>vejamos</i>	<i>cabemos</i>	<i>caibamos</i>
2PL	<i>vindes</i>	<i>venhais</i>	<i>vedes</i>	<i>vejais</i>	<i>cabeis</i>	<i>caibais</i>
3PL	<i>vêm</i>	<i>venham</i>	<i>vêem</i>	<i>vejam</i>	<i>cabem</i>	<i>caibam</i>

These allomorphs and their distribution are phonologically predictable given the effects of desinential *yod* and therefore represent old morphological alternations.²⁸ Old Portuguese also had velar allomorphy in the L-pattern from the class of Latin

28. Thus /k/ + [j] is attested in FACIŌ > *faço*, and also the old form *iaço* < IACEŌ; /t/ + [j] in METIŌ > *meço*; /l/ + [j] in *valho* < VALEŌ; /n/ + [j] in *tenho* < TENEŌ; /d/ + [j] in *vejo* < VIDEŌ; /p/ + [j] in CAPIŌ > *caibo*. Old Spanish contains remnants of the latter development only for /p/ + [j] in CAPIŌ > *quepo* and /d/ + [j] (*veya* < VIDEAM, *seya* < SEDEAM, *oya* < AUDIAM, *riya* < RĪDEAM) and /g/ + [j] (*fuya* < FUGIAM).

ingressive verbs of the type in Table 6: Old Portuguese *pareasco*, *pareasca*, *conhesco*, *conhesca*. These verbs, however, were levelled in favour of the majority root of the paradigm which lacked the velar: *pareasco* > *pareço*, *conhesco* > *conheço* under the influence of *parece* and *conhece*.

Portuguese speakers could have analysed these verbs either as having the morphological structure root + velar endings or having a different root allomorph from the rest of the paradigm which only differed in one formative, a velar consonant. What I propose is that both these analyses contrasted with other L-pattern forms which clearly showed different radical types of root allomorphy (see Table 19). Such a preference towards radical allomorphy in Old Portuguese may have encouraged speakers to reject the morphological type displayed by the L-pattern allomorphy of ingressive verbs *cognosco* and *pareasco* and thus merely regularised the root *pareasco* > *pareço* under the influence of *parece*. Thus, in Old Portuguese there could have been different morphological models/types for L-pattern forms and these were reduced to one: radical root allomorphy.

Support for this hypothesis comes from the development of the verb *poder* 'be able' in Portuguese. In Latin, this verb displayed two root allomorphs POSS- and POT- which, as shown in Table 20, a partial paradigm of the Latin verb POSSE 'be able', had a rather random distribution which was shared by no other verbs in the language.

Table 20. Root allomorphs of POSSE

	Pres. Indic.	Imperf. Indic.	Pluperf. Indic.
1SG	POSSUM	POTERAM	POTUERAM
2SG	POTES	POTERĀS	POTUERĀS
3SG	POTEST	POTERAT	POTUERAT
1PL	POSSUMUS	POTERĀMUS	POTUERĀMUS
2PL	POTESTIS	POTERĀTIS	POTUERĀTIS
3PL	POSSUNT	POTERANT	POTUERANT
	Pres. Subj.	Imperf. Subj.	Pluperf. Subj.
1SG	POSSIM	POSSEM	POTUISSEM
2SG	POSSĪS	POSSĒS	POTUISSĒS
3SG	POSSIT	POSSET	POTUISSET
1PL	POSSĪMUS	POSSĒMUS	POTUISSĒMUS
2PL	POSSĪTIS	POSSĒTIS	POTUISSĒTIS
3PL	POSSINT	POSSENT	POTUISSENT
	Pres. Inf.	Perf. Inf.	Participle
	POSSE	POTUISSE	POTĒNS

In Modern Portuguese this original distribution is lost and in its place the suppletive allomorph *poss-* is distributed according to the L-pattern as demonstrated in the present tense forms of the Modern Portuguese verb *poder* in Table 21.²⁹

Table 21. The L-pattern in Portuguese *poder*

	Pres. Indic.	Pres. Subj.
1SG	<i>posso</i>	<i>possa</i>
2SG	<i>podes</i>	<i>possas</i>
3SG	<i>pode</i>	<i>possa</i>
1PL	<i>podemos</i>	<i>possamos</i>
2PL	<i>podeis</i>	<i>possais</i>
3PL	<i>podem</i>	<i>possam</i>

This development of incursive suppletion (used in the sense of Corbett, 2007) in the L-pattern is exclusive to Portuguese, as opposed to the incursive suppletion in the N-pattern with the verb ‘go’ attested in Portuguese, Old Spanish (see footnote 21) and Catalan (see Table 12 and footnote 21) and suggests, to my mind, a preference for L-pattern allomorphy and that this allomorphy be root allomorphy which can be radically different to the other allomorphy for the lexeme.³⁰

7.1 Conclusion

In this article, I have argued for the recognition of autonomous morphological structures, or morphemes, in the Ibero-Romance verb and variation as to the morphological parsing of velar allomorphy. With regard to morphemes, I have presented data which prompt the conclusion that the allomorphy in these cells not only tends to converge upon one particular phonological form (Maiden, 2004) but also tends to converge upon a particular morphological structure or type. An appreciation of this quality of morphemes, together with an acknowledgement that velar allomorphy can correspond to root allomorphy or desinential allomorphy can facilitate more convincing explanations of instances of change in the class of velar verbs in Ibero-Romance, especially when this allomorphy coincides with diphthongisation.

29. Similar developments are also attested in other Romance languages (Old Tuscan, Cascinagrossa (Piedmont), Rumantsch (Savognin)).

30. Such preferences of Portuguese, as compared to Spanish, can, to my mind, be traced back to the different treatments of desinential *yod* within the two languages and, specifically, in *-er* verbs and *-ir* verbs (see O’Neill, 2012).

References

- Anderson, S. R. (2008). Phonologically conditioned allomorphy in the morphology of Surmiran (Rumantsch). *Word Structure*, 1, 109–134. doi:10.3366/E1750124508000184
- Anderson, S. R. (2010). Failing one's obligations: Defectiveness in Rumantsch reflexes of DÉBÈRE. In M. Baerman, G. G. Corbett & D. Brown (Eds.), *Defective paradigms: Missing forms and what they tell us* (pp. 19–34). Oxford: British Academy/Oxford University Press. doi:10.5871/bacad/9780197264607.003.0002
- Anderson, S. R. (2011). Stress-conditioned allomorphy in Surmiran (Rumantsch). In M. Maiden, J. C. Smith, M. Goldbach & M. -O. Hinzelin (Eds.), *Morphological Autonomy: Perspectives from Romance inflectional morphology* (pp. 13–35). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199589982.003.0002
- Aronoff, M. (1994). *Morphology by itself*. Cambridge, MA: The MIT Press.
- Aski, J. M. (1995). Verbal suppletion: An analysis of Italian, French and Spanish *to go*. *Linguistics*, 33, 403–432. doi:10.1515/ling.1995.33.3.403
- Badia, A. M. (1994). *Gramàtica de la llengua catalana*. Barcelona: Enciclopèdia Catalana.
- Barcos, M. Á. (2007). *El aragonés ansotano: Estudio lingüístico de Ansó y Fago*. Zaragoza: Gara d'Edicions.
- Bermúdez-Otero, R. (ms.). *The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features*.
- Black, M., & Chiat, S. (1981). Psycholinguistics without 'psychological reality'. *Linguistics*, 19, 37–61. doi:10.1515/ling.1981.19.1-2.37
- Bosque, I., & Pérez Fernández, M. (1987). *Diccionario inverso de la lengua española*. Madrid: Gredos.
- Burzio, L. (2004). Paradigmatic and syntagmatic relations in Italian verbal inflection. In J. Auger, J. C. Clements & B. Vance (Eds.), *Contemporary approaches to Romance linguistics: Selected papers from the 33rd Linguistic Symposium on Romance Languages (LSRL), Bloomington, Indiana, April 2003* (pp. 17–44). Amsterdam: John Benjamins. doi:10.1075/cilt.258.02bur
- Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: John Benjamins. doi:10.1075/tsl.9
- Bybee, J., & Pardo, E. (1981). On lexical and morphological conditioning of alternations: A nonce-probe experiment with Spanish verbs. *Linguistics*, 19, 937–968. doi:10.1515/ling.1981.19.9-10.937
- Carreira, M. (1991). The alternating diphthongs of Spanish: A paradox resolved. In H. Campos & F. Martínez-Gil (Eds.), *Current studies in Spanish linguistics* (pp. 407–445). Washington, DC: Georgetown University Press.
- Corbett, G. G. (2007). Canonical typology, suppletion and possible words. *Language*, 87, 8–42. doi:10.1353/lan.2007.0006
- Díaz González, O. J. (1986). *El habla de Candamo: Aspectos morfosintácticos y vocabulario*. Oviedo: Universidad de Oviedo.
- Eddington, D. (2004). *Spanish phonology and morphology: Experimental and quantitative perspectives*. Amsterdam: John Benjamins. doi:10.1075/sfsl.53
- García-Bellido, P. (1986). Lexical diphthongization and high-mid alternations in Spanish: An autosegmental account. *Linguistic Analysis*, 16, 61–92.
- Goyvaerts, D. L. (1978). *Aspects of post-SPE Phonology*. Ghent: E. Story-Scientia.
- Haensch, G. (2003). *Las hablas de la Alta Ribagorza*. Zaragoza: Instituto de Fernando el Católico.

- Harris, J. (1969). *Spanish phonology*. Cambridge, MA: The MIT Press.
- Harris, J. (1977). Remarks on diphthongization in Spanish. *Lingua*, 41, 261–305.
doi:10.1016/0024-3841(77)90082-1
- Harris, J. (1978). Two theories of non-automatic morphophonological alternations. *Language*, 54, 41–60. doi:10.2307/412998
- Harris, J. (1985). Spanish diphthongization and stress: A paradox resolved. *Phonology Yearbook*, 2, 31–45. doi:10.1017/S0952675700000373
- Higginbotham, J. (1991). Remarks on the metaphysics of linguistics. *Linguistics and Philosophy*, 14, 555–566. doi:10.1007/BF00632597
- Hualde, J. I. (1992). *Catalan*. London: Routledge.
- Lass, R. (1976). On generative taxonomy and whether formalisms ‘explain’. *Studia Linguistica*, 30, 139–154. doi:10.1111/j.1467-9582.1976.tb00639.x
- Lloyd, P. (1987). *From Latin to Spanish. Vol. I: Historical phonology and morphology of the Spanish language*. Philadelphia, PA: American Philosophical Society.
- Lozano Sierra, C., & Saludas Bernad, A. L. (2007). *Aspectos morfosintácticos del Belsetán: Aragonés del valle de Bielsa*. Zaragoza: Gara d’Edicions.
- Malkiel, Y. (1971). Derivational transparency as an occasional co-determinant of sound change: A new casual ingredient in the distribution of -ç- and -z- in ancient Hispano-Romance. *Romance Philology*, 25, 1–52.
- Malkiel, Y. (1974). New problems in Romance interfixation. *Romance Philology*, 27, 304–355.
- Malkiel, Y. (1993). The problem of the Old Spanish sibilants: Three consecutive new-style explanations. *Bulletin of Hispanic Studies*, 70(2), 201–211. doi:10.1080/1475382932000370201
- Maiden, M. (1992). Irregularity as a determinant of morphological change. *Journal of Linguistics*, 28, 285–312. doi:10.1017/S0022226700015231
- Maiden, M. (2001). Di nuovo sulle alternanze velari nel verbo italiano e spagnolo. *Cuadernos de filología italiana*, 8, 39–61.
- Maiden, M. (2004). Morphological Autonomy and diachrony. *Yearbook of Morphology 2004*, 137–175.
- Maiden, M. (2009). From pure phonology to pure morphology: The reshaping of the Romance verb. *Recherches linguistiques de Vincennes*, 38, 45–82. doi:10.4000/rlv.1765
- Maiden, M. (2011). Morphemes and ‘stress- conditioned allomorphy’ in Romansch. In M. Maiden, J. C. Smith, M. Goldbach & M. -O. Hinzelin (Eds.), *Morphological Autonomy: Perspectives from Romance inflectional morphology* (pp. 36–51). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199589982.003.0003
- Martín Vegas, R. A. (2007). *Morfofonología histórica del español: Estudio de las alternancias /jé/ - /e/, /wé/ - /o/ y /Ø/ - /g/*. Munich: Lincom.
- Mascaró, J. (1983). *La fonologia catalana i el cicle fonològic*. Bellaterra: Publicacions de la Universitat Autònoma de Barcelona.
- Menéndez Pidal, R. (1941). *Manual de gramática histórica española*. Madrid: Espasa-Calpe.
- Morin, Y. Ch. (1988). Morphological conditioning in phonologically transparent processes: Evidence from the evolution of vowel reduction in Vinzelles Occitan. *Canadian Journal of Linguistics*, 33, 431–442.
- Ohala, J. J. (1990). There is no interface between phonology and phonetics: A personal view. *Journal of Phonetics*, 18, 153–171.
- O’Neill, P. (2011a). *The Ibero-Romance verb: Allomorphy and the notion of the morpheme*. (Unpublished doctoral dissertation). University of Oxford.

- O'Neill, P. (2011b). Alternancias vocálicas en el presente de los verbos en *-ir*. Un análisis desde la Morfología Autónoma. *Revista de la Historia de la Lengua Española*, 6, 87–130.
- O'Neill, P. (2012). New perspectives on the effects of *yod* in Ibero-Romance. *Bulletin of Spanish Studies*, 89(5), 665–697. doi:10.1080/14753820.2012.696175
- O'Neill, P. (2015). The origin and spread of velar allomorphy in the Spanish verb: A morphomic approach. *Bulletin of Hispanic Studies*, 92(5), 489–518. doi:10.3828/bhs.2015.29
- Penny, R. (1970). *El habla pasiega: Ensayo de dialectología montañesa*. London: Tamesis.
- Penny, R. (2002). *A history of the Spanish language* (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511992827
- Perea, M. P. (2002). Flexió verbal regular. In J. Solà, M. -R. Lloret, J. Mascaró & M. Pérez-Saldanya (Eds.), *Gramàtica del català contemporani* (pp. 583–646). Barcelona: Empúries.
- Pérez-Saldanya, M. (1995). Un problema de morfología verbal románica: Les formes verbals velaritzades. In M. T. Echenique, M. Aleza & M. J. Martínez (Eds.), *Actas del I Congreso de Historia de la lengua española en América y España* (pp. 409–420). Valencia: Tirant lo Blanch.
- Pérez-Saldanya, M. (1998). *Del llatí al català: morfosintaxi verbal històrica*. Valencia: Universitat de València.
- Pirrelli, V. (2000). *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*. Pisa: Istituti Editoriali e Poligrafici Internazionali.
- Pirrelli, V., & Battista, M. (2000). The paradigmatic dimension of stem allomorphy in Italian verb inflection. *Rivista di linguistica*, 12(2), 307–380.
- Roca, I. (1990). Morphology and verbal stress in Spanish. *Probus*, 2(3), 321–350. doi:10.1515/prbs.1990.2.3.321
- Roca Pons, J. (1968). Morfologia verbal catalana. *Estudis Romànics*, XII, 227–254.
- Sampson, G. (2001). *Empirical linguistics*. London: Continuum.
- Saura Rami, J. A. (2003). *Elementos de fonética y morfosintaxis benasquesas*. Zaragoza: Garza d'Edicions.
- Schuldberg, H. K. (1984). Diphthongization in Spanish verbs. *Hispanic Linguistics*, 1, 215–228.
- Skousen, R. (1989). *Analogical modeling of language*. Dordrecht: Kluwer Academic Press.
- St. Clair, R., & Park, C. (1974). The irregular present tense verbs of Spanish. *Linguistics*, 12, 73–100. doi:10.1515/ling.1974.12.135.73
- Viaplana, J. (1984). La flexió verbal regular del valencià. In *Miscel·lània Sanchis Guarner* (pp. 391–407). Valencia: Universitat de València.
- Williams, E. B. (1962). *From Latin to Portuguese. Historical phonology and morphology of the Portuguese language* (2nd ed.). Philadelphia, PA: University of Pennsylvania Press.
- Wheeler, M. W. (1993). Changing inflection: Verbs in North West Catalan. In D. Mackenzie & I. Michael (Eds.), *Hispanic linguistic studies in honour of F. W. Hodcroft* (pp. 171–206). Llangrannog, Wales: Dolphin.
- Wheeler, M. W. (2011). The evolution of a morpheme in Catalan verb inflection. In M. Maiden, J. C. Smith, M. Goldbach & M.-O. Hinzelin (Eds.), *Morphological Autonomy: Perspectives from Romance inflectional morphology* (pp. 183–209). Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199589982.003.0010

Appendix

Lexemes which display voiceless velar allomorphy according to the L-pattern (derivatives not included). Note that this list has been compiled on the basis of Bosque and Pérez Fernández's reverse Spanish dictionary (1987). Verbs were excluded if they were catalogued in the online dictionary of the Real Academia Española as in disuse or antiquated and were not given a model of conjugation, e.g., *adhierecer*.

155 in total

placer, nacer, pacer, yacer, acaecer, escaecer, embebecer, embobecer, embarbecer, herbecer, padecer, agradecer, obedecer, humedecer, empecer, palidecer, aridecer, languidecer, lividecer, encandecer, ablandecer, resplandecer, engrandecer, ensandecer, enardecer, atardecer, verdecer, engordecer, ensordecer, enzurdecer, enmudecer, recrudecer, enrudecer, anochecer, mohecer, entibiecer, embermejecer, vejecer, hojecer, enrojecer, calecer, escalecer, enmalecer, enralecer, fortalecer, prevalecer, convalecer, establecer, ennoblecer, encrudelecer, encruelecer, envilecer, encallecer, fallecer, tallecer, embellecer, amarillecer, amollecere, pimpollecere, enorgullecer, entullecere, arbolecer, herbolecer, adolecer, entumecer, adormecer, emplumecer, encanecer, amanecer, hermanecer, permanecer, desvanecer, lozanecer, fenecer, pertenecer, juvenecer, orinecer, enruinecer, onecer, adonecer, embarnecer, encarnecer, escarnecer, ensarnecer, guarnecer, enternecer, fornecer, empequeñecer, retoñecer, empecer, entorpecer, carecer, aclarer, amarecer, parecer, aparecer, enrarecer, guarecer, entenebrecer, ensombrecer, empobrecer, crecer, podreecer, merecer, perecer, ofreecer, magreecer, negreecer, enmugreecer, alborecer, arborecer, encorecer, fosforecer, florecer, amorecer, esmorecer, favorecer, empavorecer, enfervorecer, aterecer, aborreecer, enlustreecer, obscurecer, escurecer, oscurecer, escurecer, endurecer, enfurecer, apeteecer, enaltecer, endentecer, lentecer, enllentecer, acontecer, entontecer, enfortecer, amortecer, abastecer, plastecer, entestecer, entristecer, robustecer, emputecer, embrutecer, frutecer, colicuecer, envaguecer, enceguecer, lobrequecer, enflaquecer, enriquecer, blanquecer, enfranquecer, enronquecer, encloquecer, enloquecer, embosquecer, abravecer, engravecer, altivecer, encalvecer, emplebeyecer, conocer, pubescer, evanescer, fosforescer.

The history of concatenative compounds in Spanish

María Irene Moyna

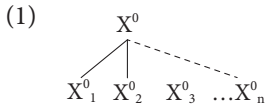
Texas A & M University

This article analyses Spanish concatenative compounds, whose constituents are of the same lexical category and structural hierarchy (e.g., *rojinegro* 'red-black'), focusing in particular on the evolution of their formal features. The first issue considered is the relative frequency of various structural configurations, namely, overt coordination ([XconjX]), linking vowel *-i-* ([XiX]), and juxtaposition ([XX]). The second issue is whether the linking vowel *-i-* is etymologically related to coordination. I propose that linking vowels can better be analysed as stem markers, distinguishing full lexemes from combining stems. Support for this proposal comes from the fact that the linking vowel was attested earlier in head-final compounds (*barbirrapado* 'beardless', lit. 'beard-shaven'), where it generalised before spreading to concatenatives.

Keywords: concatenative compounds; linking vowel; stem marker; binomial; Spanish; morphology

1. Introduction

Concatenative compounds, variously known as binomials, co-compounds, *dvandvas*, or copulative, coordinative, or additive compounds (Bauer, 2008; Olsen, 2001; Spencer, 1991; Wälchli, 2005; Whitney, 1941 [1879]) have a flat structure in common, whose constituents are not in a head-dependent relationship, but instead share structural and semantic headedness, represented in (1), where X stands for the same grammatical category, e.g., *rojinegro* 'red and black', lit. 'red-black'. Yet, as shown in Section 2, this large class of related compound patterns is linguistically complex: it may involve different lexical categories, distinct structural configurations and a variety of semantic relationships between constituents, all of them compatible with shared headedness.



Although concatenative compounds constitute a numerically small class in Spanish and exhibit only a fraction of the patterns available cross-linguistically, they deserve a more precise analysis than they have received until now. One problem with previous accounts is that concatenative compounds are often confused with endocentric compounds of the same surface form. For example, Casado (1992) uses the term *sustantivo en aposición*, i.e. noun in apposition, to refer to any $[N+N]_N$ compound. As a result, he groups together head-initial endocentric compounds (e.g., *hora-punta* ‘rush hour’, lit. ‘hour-peak’) and concatenative compounds (e.g., *sofá-cama* ‘sofa-bed’). For his part, Val Álvaro (1999, p. 4780) oscillates in his treatment of some of these compounds (e.g., *entrenador jugador* ‘trainer player’), classifying them both as head-initial endocentrics and as coordinative (i.e. concatenative) compounds. Even authors who distinguish concatenative from endocentric compounds sometimes fail to distinguish between their various subtypes. For example, Rainer (1993), while recognising the independent identity of $[N+N]_N$ concatenative patterns, does not distinguish their different semantic-syntactic classes. Thus, he groups together *actor-bailarín* ‘actor-dancer’, *otoño-invierno* ‘fall-winter’, *compraventa* ‘sale’, lit. ‘purchase-sale’, *falda-pantalón* ‘skort’, lit. ‘skirt-trousers’, which belong in distinct classes, as shown in Section 3. A better description of concatenative compounds establishes clearer structural and semantic boundaries between them and endocentric compounds with identical surface form, as well as among the various types of concatenatives, which in turn helps the increasing frequency of these patterns stand out (Moyna, 2011, p. 281). Sound analysis thus leads to recognition of the growing importance of this compounding pattern, a fact that may otherwise go unnoticed.

The first aim of this paper is to present formal changes that concatenative compounds have undergone in Spanish, in particular in terms of the structural expression of the relationship between constituents. It is demonstrated that while simple juxtaposition of constituents has always been available as a mechanism to create concatenative compounds (*abecé* ‘alphabet’, lit. ‘a-be-ce’ [1236]), over time it has become increasingly frequent for a linking vowel to appear between constituents (*carricoche* ‘cart’, lit. ‘cart-car’ [1605]).¹ The increased presence of

1. The year between brackets constitutes the earliest attestation available.

the vowel is accompanied by a decrease in the presence of compounds with overt coordination (*uso y costumbre* ‘customs’, lit. ‘use and custom’ [1256]).

The second aim is to establish whether overt coordination is the etymological precursor of the *-i-* linking vowel (as suggested in Baist, 1899, p. 474; Alvar & Pottier, 1983, p. 415). There are at least three reasons to support this hypothesis. The first is that in many concatenative compounds the semantic relationship between constituents is additive and thus, akin to coordination. For example, if something is *rojiverde* ‘red and green’, lit. ‘red-green’, it is both red and green in some spatial distribution. The second reason is that the linking vowel *-i-* is phonologically identical to the coordinative conjunction, represented by the grapheme *y* in Modern Spanish and realised as a vocalic segment [i], a semivowel [j], or a consonant [j], depending on context. The last reason is that, as hinted above, there is an increase in the use of the linking vowel as coordination is phased out, which suggests that the rise in one may be causally related to the decline in the other. However, as we shall see in Section 5, this hypothesis is not strongly supported empirically. As an alternative, it is proposed that the linking vowel is a marker of stem status for the first constituent, since its presence is accompanied by the absence of the word class marker (Harris, 1991). It will be shown that this vocalic segment first appeared between the constituents of head-final compounds, and it is hypothesised that from them it spread to concatenatives.

The paper is organised as follows. Section 2 reviews previous research regarding the general cross-linguistic features of concatenative compounds. It focuses on Romance and, more specifically, on the history of concatenative compounds and on the evolution of the linking vowel *-i-* from Latin. Section 3 summarises previous findings on the evolution of concatenative compounds in Spanish. Section 4 describes the methodology used in this study, including sources, data collection, classification criteria and quantification. Section 5 presents the findings, as regards the frequency of the various concatenative compounding patterns and the presence and status of linking vowels. It shows that there is no general evolution from coordination to compounding. Section 6 provides an alternative hypothesis to explain the data. Section 7 concludes the paper and suggests avenues for future research.

2. Background

2.1 Concatenative compounding in historical perspective

Although it is not the purpose of this article to discuss at length the notion of compound per se, some definitions are in order for the sake of clarity. In the remainder of this work, a compound is understood as a lexeme created by combining other lexemes or lexical stems (i.e. words belonging to the grammatical categories

of noun, adjective, verb, adverb or numeral) through the application of syntactic processes (Fabb, 1998; Dressler, 2006). The resulting lexical output has certain properties typical of words, such as internal fixity of constituents, atomicity to syntactic operations and semantic idiomaticity (cf. Moyna, 2011, pp. 24–34). It also has some features that are typical of syntactic structures, such as the possibility of recursion (*limpiaparabrisas* ‘windscreen wiper’, lit. ‘clean-stop-breezes’). Like syntactic operations, compounding is productive, which makes it possible to form new compounds through the repeated application of combinatory patterns. In this regard, compounds must be distinguished from other constructions with which they share some but not all features. For example, words such as *correveidile* ‘gossipmonger’, lit. ‘run-go-and-tell-him/her’, *metomentodo* ‘meddler’, lit. ‘I meddle-myself-in-everything’ and *tentenelaira* ‘hummingbird’, lit. ‘hold-yourself-in-the-air’ share with compounds the features of fixity, atomicity and idiomaticity. However, they are not productive, given that they are not formed by the application of a pre-existing syntactic pattern, but by the haphazard chunking and lexicalisation of full-fledged sentences (for more details on these *syntactic freezes*, see Miller, 1993, p. 93; Moyna, 2011, pp. 35–36; Torres Cacoullos & Bauman, in this volume, for the role of chunking in the creation of the preposition *para*).²

The existence of compounds made up of lexical constituents combined through a non-hierarchical structure is mentioned in most morphological compendia and handbooks, including Booij (2007), Fabb (1998), Katamba (1993), Haspelmath and Sims (2010) and Spencer (1991). However, there are differences in the way these compounds are identified, defined and classified. As noted in Bauer (2008), part of the existing confusion is due to the application of terminology such as *dvandva*, originally created for one language (Sanskrit), to describe the patterns of word formation in others. To clarify the issue, Bauer presents a taxonomy that distinguishes all possible semantic nuances among compounds that share the formal characteristic of combining constituents in non-hierarchical structures. Not all those possibilities are exploited by all languages. For example,

2. This account differs, then, from those that consider a compound to be any combination of words, without regard to their lexical or functional status (Maiden, 2008; Maiden et al., 2009). This difference in approach has analytical consequences. For example, in this and other works I have proposed that the verbal constituent in $[V+V]_N$ and $[V+N]_N$ compounds is a stem, at most accompanied by its theme vowel, but devoid of any tense, mood or person specifications (see also Ferrari-Bridgers, 2005, for a similar proposal for Italian). In this view, it makes no sense to discuss whether these verbs are in the third person singular or in the second person imperative, as has been traditional in the literature on deverbal Romance compounds (see Darmesteter, 1967, pp. 169–204 for an early example), since by definition they cannot be in either.

compounds with plural or collective referents (also known as co-compounds, e.g., *candrā-dityā-u* ‘the moon and the sun’, lit. ‘moon-sun-DUAL’) are most frequent in continental Eastern and South East Asia and decrease as we move west (Wälchli, 2005, p. 196).

Indeed, among the historical Indo-European languages, Sanskrit exhibits more varied nominal concatenative compounds than either Greek or Latin, as shown in (2a): it has compounds made up of two nouns with plural, collective or alternative reference. Greek presents collective compounds as well as ‘compromise’ forms, with a single referent combining features of both constituents (Bauer, 2008, p. 10). These hybrid patterns also appear in Latin, though very infrequently (Fruyt, 2002, p. 263), historically late and often modelled on Greek (Bader, 1962, p. 333). Adjectival concatenatives, denoting two qualities that co-occur in some combination, are less frequent in Sanskrit than their nominal counterparts. They are also found in Greek and Latin, although, again, in the latter they tend to be scarce and not very productive (2b). By contrast, [V+V] compounding is absent from historical Indo-European languages, so that its presence in the daughter languages must be interpreted as an innovation (cf. Kiparsky, 2009; Nicholas & Joseph, 2009; Ralli, 2009 for Modern Greek).

- (2) a. [N+N]_N
- | | |
|---------------------------------------|---|
| <i>ajāvāyah</i> < <i>aja#avāy+as</i> | [Sanskrit] |
| ‘goat and sheep PL’ | (Whitney, 1941 [1889], p. 485) ³ |
| <i>keśa-śmaśrú</i> | |
| ‘hair and beard NEUT SG’ | (Whitney, 1941 [1889], p. 486) |
| <i>jayaparājaya-</i> | |
| ‘victory or defeat’ | (Whitney, 1941 [1889], p. 485) |
| <i>andró-gunon</i> | [Greek] |
| ‘couple’, lit. ‘man-woman NEUT’ | (Debrunner, 1917, p. 41) |
| <i>andró-gunos</i> | |
| ‘hermaphrodite’ lit. ‘man-woman M SG’ | (Debrunner, 1917, p. 41) |
| <i>oxypiper</i> | [Latin] |
| ‘vinegar and pepper’ | (Bader, 1962, p. 336) |
| <i>porcopiscis</i> | |
| ‘dolphin’ lit. ‘pig-fish’ | (Bader, 1962, p. 335) ⁴ |

3. Sanskrit transliterations have been modified from Whitney to reflect modern conventions. I am grateful to Gary Miller for providing the correct transliterations.

4. I am following the interpretation provided in Bader (1962) here, although, as pointed out to me by Gary Miller (p. c.), the compound could have been endocentric, with *piscis* ‘fish’ as its head and the non-head constituent interpreted metaphorically to refer to the animal’s corpulence.

- b. [A+A]_A
śukla-kṛṣṇa- [Sanskrit]
 ‘light and dark’ (Whitney, 1941 [1889], p. 487)
vr̥tta-pīna-
 ‘round and plump’ (Whitney, 1941 [1889], p. 487)
chloro-mélās [Greek]
 ‘pale green – black’, ‘dark green’ (Debrunner, 1917, p. 40)
glukú-pikros
 ‘sweet-bitter’ (Debrunner, 1917, p. 40)
nigrogemmeus [Latin]
 ‘bright and somber’ (Bader, 1962, pp. 333–34)
dulcacidus
 ‘sweet and sour’ (Bader, 1962, pp. 333–34)

According to Hatcher (1951, Chapter 1), Sanskrit *dvandva* compounds of the type represented in (2a) correspond to an original ‘primitive’ Indo-European wave that left very few traces in Germanic, Greek and Latin, with the exception of numerals. In the fifth century B.C. a second new type of copulative compounds arose in Greek, with a first constituent in stem form (e.g., *λυκ-άνθρωπος* ‘luk-ánthropos’ ‘werewolf’, lit. ‘wolf-man’, viz. *λύκος* ‘lúkos’ ‘wolf’) Hatcher characterises these compounds as “artificial and fantastic” (1951, p. 6) and traces their origin to a [N+N] pattern that was semantically ambiguous between a determinative (i.e. endocentric) and a concatenative meaning. Thus, for example, if the compound *ταυρ-έλαφος* ‘taur-élaφος’ ‘ox-deer’ originally meant ‘a deer that is like an ox in some features’, it could have easily been extended to cover a hybrid species (half A, half B). These Greek hybrid compounds were adopted by Latin to name dual function objects (*tunico-pallium* ‘tunic-cloak’). They also made their way indirectly into the European vernaculars through Neo-Latin (e.g., *gallopavum*, *pavo-gallus* ‘turkey’) and directly in more short-lived native creations (e.g., *homme-femme* ‘man-woman’, a French calque based on *androgyné*).

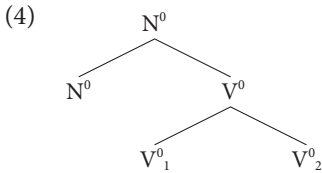
To explain the generalisation of adjectival concatenative compounds in modern European languages, Hatcher (1951, Chapter 3) identifies a third wave that originated in Classical Latin forms suffused with new meanings in Medieval Latin. For example, the Classical Latin ‘immigration pattern’ of *Gallo-Graeci* ‘Gauls who settled in Greece’, itself an adaptation of an equivalent determinative Greek pattern, prospered in Neo-Latin. From its original meaning of ‘displacement’ it came to be interpreted through the influence of German Neo-Latinists as ‘translating between or containing two languages’ (e.g., *Lexicon Graeco-Latinum* ‘Greek-Latin lexicon’ [1530]). By the 16th century, it was being used to designate something ‘belonging to two nations’ (e.g., *Gallo-Belgicus* ‘pertaining to both

France and Belgium' [1592]) and it was then adopted by European vernaculars. According to Hatcher (1951, pp. 70–79) a second pattern, which she calls “appellative” (*physico-chemical* < Neo-Latin *physico-chemicus*) can be traced back to the renewed application of Plautus’ hapax *tragicomoedia* ‘tragicomedy’, revived to describe medieval and post-medieval plays that defied classical definitions. In turn, this noun inspired a reversed adjective *comico-tragicus* [1540], created not by derivation (**tragicomoedicus*) but by compounding of the parallel adjectives. The process was spearheaded by German-speaking Neo-Latinists and the resulting pattern eventually spread to the Netherlands by the late 16th century, to Italy, England and France in the first half of the 17th century, and to Spain later in the century.

The Romance vernaculars thus followed the lead of Latin and Neo-Latin in their concatenative compounding patterns. As a consequence, nominal and adjectival concatenatives are well documented for most varieties such as Portuguese, Spanish, Catalan, French, Italian and Romanian (3a) and (3b). Concatenative compounds made up of two verbs are less frequent and unattested in Latin, but this Romance innovation is documented in Portuguese, Spanish, Catalan, and Italian (3c).

- (3) a. [N+N]_N
deputado-cantor [Portuguese]
 ‘deputy-singer’ (Alves, 1986–1987, p. 56)
chasseur-bombardier [French]
 ‘fighter plane-bomber’ (Spence, 1980, p. 85)
coliflor [Catalan]
 ‘cauliflower’, lit. ‘cabbage-flower’ (Mascaró, 1986, p. 71)
- b. [A+A]_A
dolceamaro [Italian]
 ‘bittersweet’, lit. ‘sweet-bitter’ (Scalise, 1992, p. 177)
alb-glăbui [Romanian]
 ‘white-yellow’ (Mallinson, 1986, p. 330)
- c. [V+V]_N
andirivieni [Italian]
 ‘comings and goings’, lit. ‘go-return’ (Scalise, 1992, p. 177)
[a] corre-cuita [Catalan]
 ‘in a hurry’, lit. ‘[of] run-hurry’ (Mascaró, 1986, pp. 73–74)

It must be noted that structurally, [V+V] compounds are more complex than nominal and adjectival concatenatives, since the resulting form is a nominal. This means, then, that although they appear to be flat structures that simply string together two verbs, in fact they have added layers of hierarchical structure, with a nominal empty head selecting a sequence of concatenated verbs, as in (4).



To summarise, this section has considered the possible types of concatenative compounds. It focused especially on those present in Romance, both in terms of the possible lexical category of their constituents and the semantic relationships that may obtain between them. In the next section we turn our attention to the evolution of the linking vowel *-i-* that often appears between the constituents in concatenative compounds.

2.2 The Latin origins of the linking vowel *-i-*

According to De Dardel and Zamboni (1999) the linking vowel *-i-* of Romance compounds (which they call an interfix) can be traced back to pre-Classical Latin.⁵ The authors propose that this formant is an innovation, with no antecedents in Indo-European, and that its appearance may be related to two Latin developments. The first is phonological, namely, the weakening of thematic *-o-* and *-a-* before a non-geminated consonant, which resulted in first constituents with no overt inflection (e.g., *AGR-I-COLA*). The second possibility is the generalisation of certain morphemes, such as the nominal singular genitive case or the imperative of verbs in *-i-* to compounds where they are not historically motivated. De Dardel and Zamboni (1999) find the hypothesis of a nominal development plausible because the leftmost nominal in compounds can often be interpreted as a genitive (e.g., *AGRICULTURA* ‘agriculture’, lit. ‘cultivation of the land’).

Evidence of the linking vowel can be found in numerous compound patterns present in early written Latin, including not just nominal and adjectival concatenatives (e.g., *SUOVITaurilia* ‘sacrifice of a pig, a sheep, and an ox’, *LEVIDENSIS* ‘thin, light’) but also a wide array of subordinative compounds, such as $[A+N]_A$ (e.g., *CALDICEREBRIUS* ‘hot-headed’), $[N+N]_N$ (*AQUIDUCTUS* ‘aqueduct’ which

5. A terminological clarification is in order. Whereas De Dardel and Zamboni (1999) refer to the *-i-* formant placed between compound constituents as an interfix (*interfixe*), I prefer to call it a linking vowel (*Fugenvokal*, *vocal de enlance*) and reserve the term interfix for suffixal material that appears between a lexical root and a derivational suffix or between two derivational suffixes in a sequence. Admittedly, both formants share their lack of lexical semantic content, but their structural differences are such that in my view the distinction is useful.

appears as a form to be avoided in the *Appendix Probi*, with *AQUAE DUCTUS* recommended) and $[N+V]_N$ (e.g., *SANGUISUGA* ‘leech’) (De Dardel & Zamboni, 1999, pp. 458–460). Many of these compounds are non-Classical, attested before the second century B. C. and popular in origin, as evinced by the *Appendix Probi* remark above and by their prosaic meanings. With that in mind, the authors suggest that the linking vowel *-i-* must have generalised in spoken Latin, where, unchecked by prescriptive pressures, it became an increasingly uniform overt mark of compounding.

The authors suggest that the popular trend of combining the constituents of a compound with an unstressed linking vowel *-i-* must have continued in early Romance, largely unaffected by phonetic laws. They argue that only this can account for the fact that *-i-* compounds appear in obviously popular vocabulary (names of plants, animals, tools, occupations) in varieties with little learned borrowing (Gascon, Sardinian, Corsican, Southern Italian), and in those where Latin /i/ gave way to a mid-vowel, such as Spanish. The hypothesis of popular transmission, De Dardel and Zamboni suggest, is the most general and chronologically earliest explanation for the origin of the linking vowel and, as such, it should be favoured. Yet, they emphasise that it does not preclude the possibility of alternative explanations for the presence of *-i-* in some compounds, such as later borrowings from Latin or between Romance languages.

An analysis of comparative Romance data allows De Dardel and Zamboni to posit two dialectal groupings based on the presence or absence of authentic compounding with *-i-*. The first includes Ibero-Romance, Southern Gallo-Romance, Sardinian and Italo-Romance, and corresponds spatially to areas of earlier Romanisation. In those varieties, native compounds may appear with or without linking vowels. The *-i-* appears in the popular lexicon, often in defiance of phonetic rules that lower /i/ to /e/. The second grouping includes Northern Gallo-Romance, Rhaeto-Romance and Romanian, and corresponds to geographical areas of later Romanisation. In these varieties the linking vowel only appears in late learned compounding. De Dardel and Zamboni conclude that the linking vowel corresponds to earlier patterns of compounding and only remained active over time in some varieties such as Ibero-Romance. By contrast, compounds lacking *-i-* are the result of a later tendency to remotivate compounds where the vowel was no longer interpretable as a linking morpheme.

In this section we have seen that the linking vowel *-i-* appears to be a trait of popular Latin compounding, present in many different patterns, both concatenative and endocentric. It is a good example of multiple causation in linguistic genesis, with phonetic and morphological rules conspiring to produce uniform marking of the first compound constituent. Although not present as a native feature in all Romance varieties, it is nonetheless frequent and widely spread

among the earliest areas of Romanisation, including Ibero-Romania. We now turn to the types of concatenative compounds of Spanish from a structural and historical perspective, in order to provide a precise framework for the remainder of the study.

3. Concatenative compounding in Spanish

3.1 Types of concatenative compounds

Spanish is typical of Romance in terms of the types of concatenative compounds it exhibits. There are four structural types defined by the grammatical category of the constituents: nominal concatenatives ($[N+N]_N$, (5a)), adjectival concatenatives ($[A+A]_A$, (5b)), exocentric verbal concatenatives ($[V+V]_N$, (5c)) and, finally, numeral quantifier concatenatives ($[Q+Q]_Q$, (5d)). For the purposes of this study, only the first three classes are considered.⁶

- (5) a. $[N+N]_N$
amigo-enemigo
 'friend-enemy'
compraventa
 'sale', lit. 'sale-purchase'
ajoaceite
 'type of sauce', lit. 'garlic-oil'
- b. $[A+A]_A$
económico-social
 'economic-social'
rojiverde
 'red-green'
agridulce
 'sour-sweet', lit. 'sweet and sour'

6. The type *salpimentar* 'add salt and pepper', lit. 'salt-pepper-v SUFF' could be included, since it can be analysed synchronically as formed through the affixation of verbal inflection to a concatenated nominal base, *salpimienta* 'salt + pepper' and is therefore $[N+N]_V$. The relative dating of the forms also allows for the possibility of an etymological relationship between the nominal *salpimienta* [1529] and the verb *salpimentar* [1605, in the form *se ha salpimentado*]. However, as one reviewer suggests, the adjectival *salpimentado* [1570] cannot be discarded as a possible source of the verb either. In any case, *salpimentar* and similar verb forms were not considered because there are not enough of them to constitute a major class.

- c. [V+V]_N
ganapierde
 ‘losing game’, lit. ‘win-lose’
callacuece
 ‘secretive person’, lit. ‘be silent-cook’
tira y afloja
 ‘conflict’, lit. ‘pull and loosen’
- d. [Q+Q]_Q
mil tres
 ‘one thousand and three’, lit. ‘thousand three’
treinta y cuatro
 ‘thirty-four’, lit. ‘thirty and four’
dieciséis
 ‘sixteen’, lit. ‘ten-and-six’

The semantic relationships that can be encoded by concatenative compounds vary depending on the lexical category of the constituents combined. Nominal concatenative compounds convey three basic meanings. First, they may be “co-participant compounds”, i.e. compounds that denote the two independent participants in some kind of interaction and which can only modify an external nominal: e.g., *relación madre-niño* ‘mother-child relationship’ (Bauer, 2008, p. 6). These compounds are rarely included in dictionaries and they will not be dealt with in the remainder of this study. The second type of nominal compounds are “identificational”, traditionally known as “appositional” (Bauer, 2008, p. 4; Wälchli, 2005, p. 161). In these, the relationship between constituents is intersective and the resulting compound refers to an individual that belongs simultaneously to two sets. The most common types involve professions, functions or locations: e.g., *actor-director*, *panadería-confitería* ‘bakery-pastry shop’. The third meaning conveyed by nominals is a partial blend of semantic features in a single denotation (“compromises” in Bauer, 2008, p. 10; “intermediate-denoting compounds” in Wälchli, 2005, p. 162). They often denote combinations, mixtures or new hybrid species: e.g., *gallipavo* ‘American turkey’, lit. ‘rooster-turkey’, *sureste* ‘south-east’. Finally, a less frequent meaning of nominal concatenative compounds is “alternative”, when the meanings of the two constituents are opposites: *amigo-enemigo* ‘friend-enemy’.

The adjectival concatenative compounds of Spanish are of two main kinds. The first are “co-participant compounds”, which parallel their nominal counterparts: e.g., *negociaciones franco-argentinas* ‘French-Argentine negotiations’. The second are “additive compounds” that create a predicate combining several characteristics and applying them to the same argument: *agridulce* ‘bittersweet’, lit. ‘sour-sweet’. A subtype of these additive compounds is “distributive”, i.e. the predicates apply to different sections of the argument: [*camiseta*] *albicelste* ‘white-blue [shirt]’, where

the white and the blue are applied in stripes, spots, or some other pattern (Real Academia Española, 2009, § 11.6b).

Finally, the very scarce verbal concatenative compounds of Spanish are mostly “additive”: they denote two events that occur simultaneously, intensively, or alternatively. Consider *sanamunda* ‘type of plant [*Thymelaea pubescens* L.]’, lit. ‘heal-clean’, *picapica* ‘magpie’, lit. ‘peck-peck’ and *tira y afloja* ‘conflict’, lit. ‘tighten and release’, respectively.

3.2 History of concatenative compounds in Spanish

Historical studies of native compounding in Spanish (e.g., Moyna, 2011) show that both nominal and adjectival concatenative compounds are attested from the earliest Spanish data available, as shown in the data presented in Section 2. Many adjectival and nominal compounds are inherited rather than created in Spanish, but their internal structure continues to be transparent: [A+A]_A *sacrosanto* ‘sacrosanct’, lit. ‘holy-saint’ [1046]; [N+N]_N: *usofructo* ‘usufruct’, lit. ‘use-enjoyment’ [c. 1250]. By contrast, verbal concatenative compounding appears later. It is frequent for verbal concatenative compounds to be borrowed from another language and to undergo folk etymology. For example, according to Corominas and Pascual (1980–1991) *cortapisa* ‘limitation’, lit. ‘cut-tread’ [ad 1435], formerly ‘adornment on clothes’, comes from Old Catalan *cortapisa* ‘id.’, possibly meaning ‘adornment of different material sewn on bedcovers’, in turn from Lat. CŪLCĪTA PĪNSA ‘pleated mattress’. Thus, the form is semantically opaque, but still provides formal evidence of a compounding pattern (s.v. *cortapisa*, vol. 1, p. 212).⁷

From a quantitative perspective, concatenative compounds are infrequent in Spanish overall (Moyna, 2011, p. 280): nominals constitute 5.3% of the total (ibid., p. 226), adjectivals reach 5.7% (ibid., pp. 237–238) and verbals remain below 1% (ibid., p. 245). However, relative to all compounds, their overall percentage more than doubles between the 11th and the 21st centuries, from 5.4% to 12.4%

7. The issue of the etymology of *cortapisa* is complex. As stated earlier, Corominas and Pascual are of the opinion that it originated in Catalan, whence it was borrowed into Spanish and later Portuguese (*quartapiza* from the mid-16th century). The word is well attested in the 15th century both in Spanish and in “many Catalan documents” (Alcover, 1443, J. Roig, ca. 1460, and *Tirant lo Blanch*, ad 1490) (Corominas & Pascual, 1980–1991, vol.1, p. 212). Evidence of the Catalan origin is the existence of the variant *cortapeu*, which appears in Aragonese inventories dated in 1362. On the other hand, the *Diccionari Català-Valencià-Balear* (Alcover & Borja Moll, 1993) states that the etymology of the word is uncertain. They cast doubt on the compound CULCITA + PISARE (?) as a possible origin and propose instead that “probably Catalan took *cortapisa* from Castilian, a language in which the word is very vital and has acquired secondary meanings it lacks in Catalan”.

(Figure 1). Additionally, when nominal, adjectival, and verbal patterns are considered individually, there are differences in their frequencies over time. For example, $[N+N]_N$ concatenatives peak in the 1500–1600s at 6.5% and drop to 5.5% by the 1900s. By contrast, $[A+A]_A$ concatenatives remain under 3% until the 1700s but triple their relative frequency in the following two centuries. On the other hand, $[V+V]_V$ have always been very infrequent (under 2%).

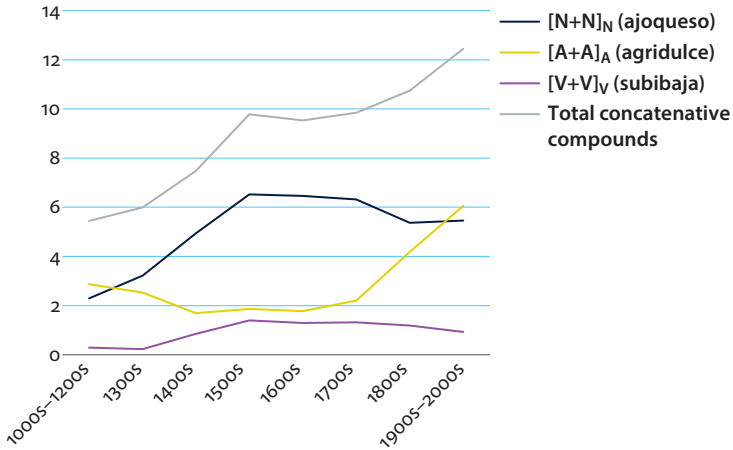


Figure 1. Frequency of concatenative compounds over time as a percentage of all compounds (from Moyna, 2011, p. 283)

The analysis in Moyna (2011, pp. 219–252) distinguishes concatenative compounds by lexical category but does not focus on the formal processes available to combine constituents. The present study focuses on the presence or absence of a linking vowel in these compounds and ascertains the possible etymological source and meaning of this element. The formal evolution of compounds of the type presented here has been implicitly or explicitly assumed to proceed through univerbation of coordinated expressions in a process of increased formal tightness: $[XconjX] > [XiX] > [XX]$ (Kastovsky, 2009, pp. 328–331; Wälchli, 2005, p. 251). The suggestion that coordination must be at the root of all non-hierarchical compounding patterns implicitly underlies many of the names given to these compounds (coordinative, copulative, conjoined). Wälchli ventures to say that sources other than coordination for these compounds are ‘minor’, although, that comment refers to co-compounds rather than to the singular denotation compounds found in Spanish. In what follows I will show that this hypothesis fails at the empirical level, and new explanations are needed to better account for the data.

4. Methodology: Sources, data collection, classification and quantification

The compounds used in this study come from the dataset of a larger-scale project that traces the origins and evolution of all major native compound types in Spanish.⁸ The list was culled from ten lexicographical sources spanning the documented history of the language between the 11th and the 21st centuries (see Moyna, 2011, pp. 67–72 and for list of dictionaries used, see *Corpus*). In all, a little under 3,600 compounds were found for all periods; each one was checked against CORDE/CREA and the online *Nuevo tesoro lexicográfico de la lengua española* to ascertain its earliest and latest attestations.

For the study reported here, all concatenative compounds with adjectival, nominal or verbal constituents found in Moyna (2011) were subjected to further testing to determine their structural variability. For that, each compound was searched again in CORDE to determine all possible orthographic and morphological variants ([XyX], [XiX], [X#X] and [XX]).⁹ By using the asterisk ([X*X]) it was possible to search simultaneously for the conjunction in its multiple orthographic and phonetic variants (*y*, *i*, *e*, *et* and *é*), all of which were considered equivalent for the purposes of the study.

Orthographic and morphological variants posed challenges in the search process. In order to make sure that all variants present in CORDE were found, asterisks were used instead of letters known to have alternate orthographies. Thus, for example, by searching for *a*a*eite*, it was possible to locate *ajiaceite*, *ajoaceite* and spelling variants (e.g., *agiaceite*). It was also important to make sure that morphosyntactic variants (plurals, feminine forms) were not missed; for that purpose, compounds were searched with asterisks in place of their inflectional marks. For instance, searching for *agr*dul*e** allowed for the location simultaneously of forms such as *agradulce*, *agrodulce*, *agridulce*, their variants with *c* and *ç* and any plurals of those forms present in the database. For the specific case of [XiX], the first constituent was searched with and without its inflectional ending. For example, for the compound *metisaca* ‘game’, lit. ‘put in-take out’, the strings *mete y saca*, *meteisaca* and *metisaca* were all searched.

8. For an explanation of the criteria used to determine whether a compound was native or learned, see Moyna (2011, p. 42).

9. The CREA database could not be used for the purposes of this study, because, unlike CORDE, it does not allow searches for wildcards of the type *X * X* or *X y X*, since the conjunction is taken to be the Boolean operator rather than the coordinating conjunction. The *Nuevo tesoro lexicográfico de la lengua española* is also of limited use, since one can only look up headwords, so that the types of expressions sought (e.g. [N y N] and [N#N]) are in most cases not searchable.

Once all orthographic variants were obtained from CORDE, they were grouped together according to their internal structure. For instance, juxtaposed forms (e.g., *mete saca*, *metesaca*) were counted together. Each combination of constituents was tabulated in all its attested forms ([XconjX], [XiX] and/or [XX]), and each variant was dated individually. Thus, for example, the combination of the two nouns *ajo* ‘garlic’ and *queso* ‘cheese’ was found in the three configurations, and each one was dated as follows: [XconjX] *ajo y queso* [1528], [XiX] *ajiqueso* [1589] and [XX] *ajoqueso* [1656].

The first objective of this study was to count the relative frequency of use of each formal pattern. For this purpose, compounds formed with the same constituents but with different internal structure were counted and dated independently. The second objective was to determine whether a general historical sequence could be proposed for the different patterns. To test this, compounds formed with the same constituents and meaning but different formal processes were compared to each other in terms of their first attestations. This made it possible to establish whether variants had appeared simultaneously or in different historical periods. In all cases, quantitative data were subjected to Fischer’s exact test in order to ascertain statistical significance under the assumption that the data obtained were a random sample of the actual number of concatenative compounds present in Spanish. All data manipulation and statistical analysis was performed in Stata MP.¹⁰

Before we move on to the findings, one final note is in order regarding [XconjX] forms. In the course of the search process, constituents in a compound with the structure [XiX] or [XX] were often also found in coordinated phrases with the structure [XconjX]. However, not all of those combinations were counted as compound variants, because they did not exhibit the formal features of compounding (atomicity, inseparability, fixity) or the semantic hallmarks of lexemes (idiomaticity, non-compositionality). Therefore, it was more reasonable to classify them as syntactic coordination. Take, for example, the contrast between the [XconjX] structure in (6a) and the [XiX] compound in (6b).

- (6) a. *Delante del buhío del cacique estaban debajo de una barbacoa hasta veinte indios, pintados de bixa e de xagua, que es rojo e negro, e con muchos e lindos penachos, cantando de pie*

‘In front of the chief’s hut there were up to twenty Indians under a barbecue shed, painted in *bixa* and *xagua*, which is red and black, and with many pretty headdresses, singing where they stood’

(1535–1557, Gonzalo Fernández de Oviedo,
Historia general y natural de las Indias, CORDE)

10. The software that was used is StataCorp LP. Stata MP, version 11.2.

- b. *Miró en las margaritas, mártires diezmadas por la rueda y el casco; en las rojinegras amapolas; en los narcisos, que guardan oro entre la nieve*
 'He looked at the daisies, martyrs decimated by wheel and hoof; at the red-and-black poppies; at the daffodils, which hold gold in the snow'
 (1910, José Enrique Rodó, *Motivos de Proteo*, CORDE)

In both cases the adjectival constituents *rojo* 'red' and *negro* 'black' are joined in a single predicate. However, only the second example exhibits idiomaticity, a typically lexical feature, since the predicates are applied to the argument distributively. That is, each poppy in (6b) is red and black in parts, whereas in (6a) the colors can be applied to each Indian individually or to the entire class (so that some were painted red, others black, and others both). In doubtful cases like (6a) the [XconjX] sequence was not counted as a compound.

5. Findings

This section presents the quantification of concatenative compound frequencies. Section 5.1 first considers the overall preference of nominal, adjectival and verbal compounds for the three possible structural patterns presented (i.e. [XX], [XiX] and [XconjX]). Then it shows the evolution of these preferences over time. Section 5.2 considers whether it is common for concatenative compounds formed with the same two constituents to exhibit formal variability. Finally, Section 5.3 isolates those variable concatenative compounds in order to establish whether the emergence of [XconjX], [XiX] and [XX] variants is historically staggered in a discernible chronological sequence.

5.1 Frequency of structural patterns in Spanish concatenative compounds

The relative frequency of the various compounding patterns available to form concatenative compounds in Spanish provides evidence that all three configurations have been used to create compounds in all lexical categories (Table 1). The most frequent pattern overall was simple juxtaposition [XX] (53.8%), followed by the pattern with overt coordination [XconjX] (25.3%) and, finally, the pattern with an *-i-* linking vowel [XiX] (20.9%). However, these preferences were not uniform across lexical categories. Specifically, verbal and adjectival compounds favoured juxtaposition by a statistically significant margin. Nominal compounds exhibited a less marked preference for this pattern. Although overt coordination was not preferred by any lexical category, it was more tolerated by nominal compounds (33.3%) than by either the adjectival or the verbal patterns (16.1% and 18.2%, respectively).

Table 1. Concatenative compounding patterns by lexical category
(Fisher's exact test, $p = 0.003$)

	[XconjX]	[XiX]	[XX]	Totals
Nominal	68 (33.3)	37 (18.1)	99 (48.5)	204 (100)
Adjectival	25 (16.1)	39 (25.2)	91 (58.7)	155 (100)
Verbal	6 (18.2)	6 (18.2)	21 (63.6)	33 (100)
Totals	99 (25.3)	82 (20.9)	211 (53.8)	392 (100)

The frequency of each concatenative compound pattern over time revealed that juxtaposition of constituents has always been very popular, accounting for between one third and two thirds of all new compounds in a given period (Table 2). By contrast, there were noticeable shifts in the frequency of compounding with overt coordination and with linking vowels. Whereas overt coordination was present in almost 40% of the new compounds in earlier periods, it tapered off over time and fell to under 15% in the last three centuries. By contrast, new compounds formed with linking vowels exhibited the reverse tendency. Their early frequency was below 15% but doubled between the first and second period considered.

Table 2. Concatenative compounding patterns over time (Fisher's exact test, $p < 0.0001$)

	[XconjX]	[XiX]	[XX]	Totals
1000s–1400s	20 (32.3)	8 (12.9)	34 (54.8)	62 (100)
1500s–1700s	53 (38.4)	35 (25.4)	50 (36.2)	138 (100)
1800s–2000s	26 (13.5)	39 (20.3)	127 (66.1)	192 (100)
Totals	99 (25.3)	82 (20.9)	211 (53.8)	392 (100)

5.2 Structural variability of concatenative compounds

Concatenative compounds can exhibit structural variability; in other words, the same constituents may appear in different configurations, joined by a coordinating conjunction [XconjX], by a linking vowel [XiX] or simply juxtaposed [XX]. In principle, the same two constituents may appear in all three combinations, in two or only in one. For example, the constituents *quita* 'remove' and *pon* 'put' appear in all three forms: [*de* *quita y pon* 'removable' [1597–1645], [*de* *quitaipón* [1926] and *quitaipón* 'removable ornament' [1839]. The constituents *gira* 'turn' and *pliega* 'fold' appear in two forms: *giripliega* 'type of laxative' [c. 1381–1418] and *girapliega*

[c. 1471].¹¹ Yet, as shown in Table 3, typically, compounded constituents appear consistently in a single configuration. Around 70% (or 208) of concatenative compounds have a stable form, a rate that exhibits little variation by lexical category. One fourth of compounds appear in two forms, and a little under 5% exhibit all possible configurations, with few differences by lexical category. Even allowing for possible gaps in the data, this low percentage suggests that the vast majority of concatenative compounds was formed directly by combining constituents through one of the three patterns, rather than by remodelling one compound pattern to fit another.¹²

Table 3. Concatenative compound configurations by lexical category

	Adjectival	Nominal	Verbal	Totals
[XconjX], [XiX], [XX]	3 (2.6)	5 (3.2)	2 (8.7)	10 (3.4)
[XconjX], [XiX]	4 (3.4)	9 (5.8)	2 (8.7)	15 (5.1)
[XconjX], [XX]	14 (12.1)	27 (17.3)	2 (8.7)	43 (14.6)
[XiX], [XX]	15 (12.9)	2 (1.3)	2 (8.7)	19 (6.4)
[XconjX]	4 (3.4)	27 (17.3)	0 (0.0)	31 (10.5)
[XiX]	17 (14.7)	21 (13.5)	0 (0.0)	38 (12.9)
[XX]	59 (50.9)	65 (41.7)	15 (65.2)	139 (47.1)
Totals	116 (100)	156 (100)	23 (100)	295 (100)

5.3 Historical relationship between concatenative compound variants

The last issue is the historical relationship between the variant forms of a given compound. Under the assumption of an evolution that would proceed by the gradual stripping of overt coordination (univerbation or agglutination, Fruyt, 1990), the juxtaposed pattern should be the endpoint: [XconjX] > [XiX] > [XX]. Therefore, the first step is to ascertain whether forms with linking vowels can be systematically traced back to those with overt coordination. The second step is to determine whether there is a historical precedence between the [XiX] pattern

11. The lexeme *girapliega* did not originate as a compound but rather was remodeled through folk etymology using the verbal pattern. Corominas and Pascual suggest the following etymology: *jirapliega* (also *geripliega* in Percivale, 1591) < Gr. *ιερὰ + πικρά* ‘ierá + pikrá’ ‘saint bitter’ via L.Lat. GIRAPIGRA (Corominas & Pascual, 1980–1991, vol. 3, p. 506, s.v. *jerarquía*). See also the numerous variants presented in Herrera (1996, vol. 1, p. 771, s.v. *gera pigra*).

12. Given the high number of cells, and the fact that many of them are empty, it is not possible to calculate statistical significance for this table using Fisher’s exact test.

and the juxtaposed structure [XX]. If the linking vowel were indeed derived from coordination, one would expect it to appear more frequently between constituents that were previously coordinated than between those that were previously juxtaposed.

If it can be shown that in most compounds with [XconjX] and [XiX] configurations the former precedes the latter historically, then this supports the hypothesis of a coordination origin for the linking vowel. However, close examination of the data provides only partial support for this hypothesis and suggests that the path from conjunction *y* to linking vowel *-i-* was not generalised, but highly dependent on the lexical category of the compound. Whereas most adjectival and verbal compounds of the [XiX] structure have conjoined antecedents, most nominal compounds (64.3%) do not (see Hatcher, 1951, pp. 22–23, footnote 46 for a hypothesis that anticipates these findings).¹³ Since nominal compounds are prevalent overall, when all compounds are considered together, only about half (13 out of 25) present the attestations in a chronological order compatible with the [XconjX] > [XiX] evolution. In the remaining 12 cases, the two forms are attested less than a hundred years apart or show the reverse chronological order, with [XiX] preceding [XconjX] by a hundred years or more (Table 4).¹⁴ Additionally, as shown by Fisher's exact test, the differences between compound categories are not statistically significant.

Table 4. Concatenative compounds with [XconjX] and [XiX] forms in relative order of first attestation (Fisher's exact test, $p = 0.2$)

	[XconjX] before [XiX]	[XconjX] after/together with [XiX]	Totals
Adjectival	5 (71.4)	2 (28.6)	7 (100)
Nominal	5 (35.7)	9 (64.3)	14 (100)
Verbal	3 (75.0)	1 (25.0)	4 (100)
Total	13 (52.0)	12 (48.0)	25 (100)

13. One anonymous reviewer suggests that the fact that verbal concatenative compounds tend to have conjoined antecedents more frequently may be linked to other distinguishing features of these compounds such as their more complex structure and their later appearance. However, the fact that verbals do not seem to differ from adjectivals in terms of their possible conjoined antecedents weakens this hypothesis. It thus seems more reasonable to include verbal compounds together with the remaining lexical categories as a way to establish a three-way comparison.

14. Given the possible gaps in documentary evidence, it seemed that a difference of one hundred years in attestation dates was necessary to provide robust evidence of an actual time lapse.

In cases where the [XconjX] > [XiX] sequence is justified, it is possible that what started as an overt conjunction did end up as a linking vowel. For example, the first attestations of some adjectival, nominal and verbal concatenative compounds offer evidence for the phrase-to-compound evolution (7a)–(7c).

- (7) a. *agraz & dulçes* [1450–1500] > *agridulce* [1576–77]
 b. *cal é canto* [1345] > *calicanto* [1554]
 c. *mete y saca* [ad 1797] > *metisaca* [1909] (Data from CORDE)

Yet, for an even larger number of compounds this historical sequence is not found (8). This suggests, then, that although [XconjX] is a possible precursor to compounds of the structure [XiX], it is by no means the only possible source.

- (8) a. *tontos y vanos* [1585] ~ *tontivano* [1599 – 1622]
 b. *capissayo* [ad 1440] > *capa y sayo* [c. 1500]
 c. *vaivén* [1427–1428] > *va y ven* [1890] (Data from CORDE)

The second historical relationship to consider is that between [XiX] compounds and their [XX] counterparts. As Table 5 shows, in the majority of these pairs, [XX] precedes [XiX] rather than the reverse. In other words, for most pairs, the juxtaposed compound appears earlier or simultaneously with the one with a linking vowel. Consider for example, *clavezínbalos* [1481–1496] and *duermevela* [1825], which precede *clavicémbalo* [1600–1713] and *duermivela* [c. 1923–1936] by over a century respectively, and *verdes bermejas* [ad 1500], which comes earlier than *verdibermejo* [1956] by over four centuries. Although the difference is not statistically significant, the evidence still suggests that those compounds were ‘retrofitted’ with a linking vowel after they were created through juxtaposition. In the process, the word class marker of the first constituent was generally dropped from the compound, leading to the presence of a first constituent stem (*clavi-*, *duermi-*, *verdi-*), rather than a full-fledged lexeme.

Table 5. Concatenative compounds with [XiX] and [XX] forms in relative order of first attestation (Fisher’s exact test, $p = 0.7$)

	[XiX] before [XX]	[XiX] after/together with [XX]	Totals
Adjectival	2 (11.1)	16 (88.9)	18 (100)
Nominal	1 (14.3)	6 (85.7)	7 (100)
Verbal	1 (25.0)	3 (75)	4 (100)
Totals	4 (13.8)	25 (86.2)	29 (100)

6. An alternative hypothesis

This section recaps and weighs the evidence in favour and against the coordination to linking vowel path and then provides an alternative to this etymology. As seen in Section 1, there are several reasons to propose a relationship between linking vowels and overt coordination, including similar phonetic and semantic properties and the fact that the linking vowel in new compounds grows over time concomitantly with a decrease in overt coordination. Indeed, as seen in (7), in several cases of variation, an earlier compound with coordination is followed by a later alternant with a linking vowel. However, the case for a general etymological path deriving the linking vowel from coordination is weak, especially for nominal compounds. First, there are 25 [XiX] concatenative compounds with a [XconjX] counterpart compared with 57 with no attested coordinated alternant. Moreover, when both forms do exist, almost half of the time (48%) the historical sequencing of first attestations is not what would be expected if the conjunction preceded the appearance of the linking vowel. In other words, [XiX] compounds do not depend on the existence of earlier coordinated variants to emerge. The creation of [XiX] compounds through insertion of *-i-* between previously juxtaposed constituents [XX] weakens the hypothesis even further.

There are two additional pieces of evidence against the coordination origin of [XiX] compounds. The first comes from the fact that in Medieval Spanish, the high frequency of *e* and *et* as orthographic variants of the coordinating conjunction suggest that there was a phonetic alternation between high and mid-front realisations ([i], [e]). It is unclear why only the first of these phonetic variants would appear in concatenative compounding. The second argument comes from the fact that a large number of concatenative compounds are joined by means other than *-i-*, such as the linking vowel *-o-*, which is impossible to derive etymologically from the coordinating conjunction. Although there are etymological differences between the two vowels, with *-o-* originally restricted to first lexical constituents of Greek origin in Neo-Latin and only later extended to purely Latinate or vernacular stems (Hatcher, 1951, pp. 55, 83), there are no observable semantic differences between these two vocalic formants, weakening the logical necessity for one of them to have overt coordination as a precursor.¹⁵

15. I wish to thank an anonymous reviewer for the first of these observations and Hiroto Ueda for the second. A careful appraisal is needed of the specific domains of *-o-* and *-i-* insertion in Spanish, together with the relative evolution of these two formants over time (see Hatcher, 1951, p. 54).

The alternative hypothesis proposed and supported here is that linking vowels in concatenative compounds result from the extended application of a linking element present in other Spanish compound patterns. These linking vowels appear in head-final compounds with nominal first constituents such as $[N+A]_A$ (e.g., *barbirrapado* ‘beardless’, lit. ‘beard-shaven’ [c. 1280]) and head-final $[N+N]_N$ (e.g., *caprifigo* ‘wild fig tree’, lit. ‘goat-fig’ [1495]). In a move reminiscent of the replacement of $[XX]$ by $[XiX]$ compounds shown in Section 5.3, in head-final compounds too, a pattern with full-fledged first constituents gave way over time to one with stems (cf. Hatcher, 1951, p. 22).

The availability of large databases allows us to quantify the evolution of the preference for one $[N+A]_A$ pattern over another. Before the 1500s, almost 30% of compounds with that structure were formed with full lexeme nominal non-heads on the left, and an additional 20% exhibited a bare stem with no linking element; after that date, the use of the linking vowel became virtually categorical (Table 6) (cf. also Munthe, 1889). This change often resulted in the remodelling of some pre-existing compounds to fit the new pattern (e.g., *bocaroto* ‘[of] soft mouth’, lit. ‘mouth broken’ [1246–1252] > *boquirroto* [1508]). Similarly, in head-final compounds with the structure $[N+N]_N$ there was a steady (albeit not as drastic) drop in the percentages of compounds that featured a full lexeme as their first constituent (Table 7). This drop was also accompanied by a replacement of full word constituents by stems with linking vowels (e.g., *gallocresta* ‘wild clary’, lit. ‘rooster-comb’ [ca. 1300] > *gallicresta* [1494]), resulting in the elimination of word-internal inflection.

Table 6. Nominal constituents in full lexeme, stem, and stem + linking vowel in $[N+A]_A$ compounds (Moyna, 2011, p. 144) (Fisher’s exact test, $p < 0.001$)

	Full lexeme	Stem	Stem + linking vowel	Totals
1000s–1400s	8 (29.6)	6 (22.2)	13 (48.2)	27 (100)
1500s–1600s	0 (0)	5 (4.2)	113 (95.8)	118 (100)
1700s–1900s	5 (3.5)	10 (7.1)	127 (89.4)	142 (100)
Totals	13 (4.5)	21 (7.3)	253 (88.2)	287 (100)

Table 7. Nominal in full lexeme, stem, and stem + linking vowel in $[N+N]_N$ head-final compounds (Moyna, 2011, p. 180) (Fisher’s exact test, $p < 0.001$)

	Full lexeme	Stem	Stem + linking vowel	Totals
1000s–1400s	20 (50)	10 (25)	10 (25)	40 (100)
1500s–1700s	4 (14.3)	13 (46.4)	11 (39.3)	28 (100)
1800s–1900s	19 (7.3)	33 (12.7)	207 (80)	259 (100)
Totals	43 (13.2)	56 (17.1)	228 (69.7)	327 (100)

The dates of first attestation offer additional support for the hypothesis that the linking vowel started in head-final compounds and that these compounds offered a model for concatenative compounds. Firstly, linking vowels are attested earlier for head-final compounds than for concatenative compounds. Thus, for example, the earliest head-final adjectival compounds with linking vowels are attested in the 1200s: *rostrituerto* ‘[of] twisted face’, lit. ‘face-twisted’ [1240–1250], *boquimuelle* ‘[horse] of soft mouth’, lit. ‘mouth-soft’ [ca. 1275] and *barbirrapado* ‘clean-shaven’, lit. ‘beard-shaven’ [ca. 1280]. The earliest examples of *-i-* in concatenative compounds follow in the 1300s: *arquibanco* ‘bench’, lit. ‘chest-bench’ [1325] *altibaxo* ‘vicissitudes’, lit. ‘high-low’ [1330] and *giripliega* ‘laxative’, lit. ‘turn-fold’ [1381]. Secondly, the earliest head-final compounds with linking vowels, attested in the 1200s, actually precede either of the two forms of the earliest compounds that exhibit both [XX]/[XconjX] and [XiX] alternants. In fact, the [XX]/[XconjX] compounds attested during the 1200s never evolved into [XiX] forms; we have to wait until the 1300s to find compounds that eventually developed both forms. Consider, for example, *cal y canto* [1345] > *calicanto* [1554] and *arte y maña* [1499] > *artimaña* [1528]. In other words, the [XconjX]/[XX] > [XiX] shift was in fact stimulated by the prior existence of linking vowels in head-final compounds. The timeline in Figure 2 summarises this proposed chronology.

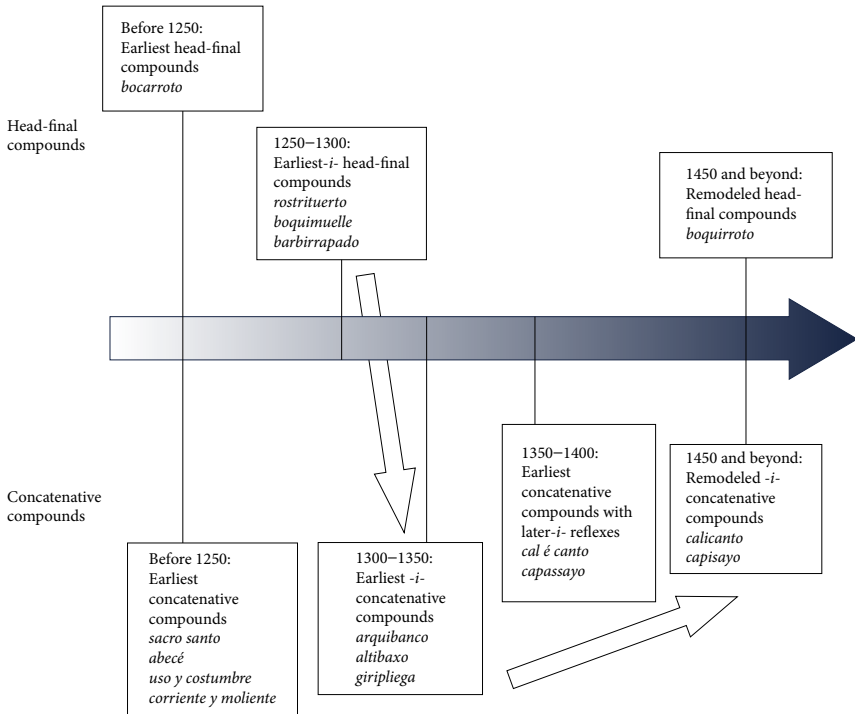


Figure 2. Timeline and hypothesised path of transmission of the linking vowel from head-final to concatenative compounds

One notable difference between the evolution of linking vowels in concatenative compounds and in head-final compounds is the failure of the vocalic segments to become categorical in the former. Whereas almost 90% of the $[N+A]_A$ compounds and 80% of the head-final $[N+N]_N$ compounds exhibit linking vowels in contemporary Spanish (Table 6 and Table 7 respectively), only 20% of concatenatives do so (Table 2). This difference may be related to the heterogeneous semantics of concatenative compounds. Whereas the syntactic-semantic relationship of constituents in $[N+A]_A$ and $[N+N]_N$ compounds is uniformly dependent-head, in concatenatives a whole range of semantic relationships is available (Section 2). The linking vowel is quite common in hybrid compounds, but it has never been an option in concatenative compounds with identificational meaning (“appositionals”). For example, *-i-* is possible in *agrodulce ~ agridulce* ‘sweet and sour’, lit. ‘sour-sweet’, but not in *compra venta* ‘sale’, lit. ‘sale-purchase’ (**compriventa*). These semantic differences may have halted the generalisation of linking vowels to all concatenative compounds.

7. Conclusions

Taken as a whole, the findings of this study provide a clear picture of the history of a compounding pattern of growing importance in Spanish and Romance more generally. To summarise, the overall frequency of concatenative compounds has increased over time in Spanish to the point that they now constitute around 12% of all compounds, divided more or less equally among nouns and adjectives, with a sprinkling of exocentric verbal compounds. Throughout history, three main structures have been used to create these compounds: juxtaposition ($[XX]$) has always been employed, whereas overt coordination ($[XconjX]$) has decreased over time, and the use of linking vowels ($[XiX]$) has increased, particularly after the 1500s.

Although in some specific cases the coordinative conjunction is very likely to have given rise to linking vowels through univerbation, particularly in the case of $[XiX]$ adjectival and verbal compounds, for most nominal compounds this etymological path cannot be adduced. It is even more improbable that $[XiX]$ compounds gave way to juxtaposed compounds by elimination of the vowel, as one would expect if the vowel were a remnant of overt coordination. In fact, the dating of allomorphs suggests that generally the juxtaposed compounds were precursors of equivalent compounds with linking vowels, rather than the reverse.¹⁶

16. Incidentally, a similar situation obtained in Sanskrit (Gary Miller, p. c.). In Vedic, *dvandvas* were distinct from other compounds because they bore two stress accents and were

It is more likely that the origin of the linking vowel is traceable to other compounding patterns in which it started to be used before it was adopted by concatenatives. In particular, head-final compounds offer a plausible model, since their linking vowel pattern generalised rapidly and reached quasi-categorical status around 1500. The fact that the constituent joined by a linking vowel was a nominal in both head-final patterns aligns well with the absence of coordinated precursors for nominal concatenatives. In the scenario proposed, the linking vowel in concatenative compounds does not have a specific coordinative meaning. This is in agreement with the restriction of linking vowels to concatenative compounds that have a hybrid meaning, i.e. those whose two constituents must be combined in a predicate within the compound itself.

The results presented here can be reinforced with comparative data from other Iberian and non-Iberian Romance languages (as suggested by De Dardel & Zamboni, 1999). It is of special interest to establish the possible formal and semantic patterns present in each variety and their evolution. The hypothesis presented in this study can be strengthened if it is shown that the presence of linking vowels in concatenative compounds is correlated with the prior or concomitant existence of similar segments in head-final patterns rather than of parallel coordinative conjunctions.

Acknowledgements

The author wishes to thank Virginia Fajt, Frances Getwick, Verónica Loureiro-Rodríguez, Gretchen Miller, Georgianne Moore, Israel Sanz-Sánchez, Gary Miller, David Pharies, and two anonymous reviewers for providing extensive and valuable feedback on this paper, and May Boggess for her help with the statistical analysis. She would also like to acknowledge the audiences of the Workshop on Ibero-Romance of the 20th International Conference on Historical Linguistics (Osaka, Japan) and of the 2011 Hispanic Linguistics Symposium (Athens, Georgia), especially Gary Baker, Miriam Bouzouita, Steve Dworkin, José Ignacio Hualde, Fernando Martínez-Gil, Enrique Pato, Diana Ranson and Hiroto Ueda for useful comments and suggestions. All errors are my own.

in the dual form (e.g. *pitārā-mātārā* ‘parents’, lit. ‘father-mother-MASC DUAL’, *mātārā-pitārā* ‘id.’, lit. ‘mother-father-MASC DUAL’). Over time, *dvandvas* started to lose these formal features, through reduction of stress and loss of the dual marking. At that point, the use of the conjunction became more frequent as a distinguishing element (e.g. *pitārā-mātārā-ca* ‘parents’ lit. ‘father-mother-CONJ’, VS 9.19) and coordinated forms started to alternate with juxtaposition (e.g. *mātaripitari* ~ *mātari-ca-pitari-ca* ‘parents’, lit. ‘mother-CONJ-father-CONJ’, Edicts of Aśoka). In other words, in Sanskrit too, juxtaposition preceded the presence of coordination in concatenative compounds.

References

- Alcover, A. M., & Borja Moll, F. de. (1993). *Diccionari català-valencià-balear: inventari lexicogràfic i etimològic de la llengua catalana en totes les seves formes literàries i dialectals*. (With the collaboration of M. Sanchis Guarner & A. Moll Marquès). Palma: Moll.
- Alvar, M., & Pottier, B. (1983). *Morfología histórica del español*. Madrid: Gredos.
- Alves, I. M. (1986–1987). Aspectos da composição nominal no português contemporâneo. *Alfa*, 30/31, 55–63.
- Bader, F. (1962). *La formation des composés nominaux du latin*. Paris: Éditions “Les Belles Lettres”. doi:10.3406/ista.1962.1010
- Baist, G. (1899). Longimanus und manilargo. *Romanische Forschungen*, 10, 471–474.
- Bauer, L. (2008). Dvandva. *Word Structure*, 1, 1–20. doi:10.3366/E1750124508000044
- Booij, G. (2007). *The grammar of words: An introduction to morphology*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199226245.001.0001
- Casado, M. (1992). Los compuestos de ‘sustantivo + sustantivo’ en aposición. Su tratamiento en la tradición lingüística española. In J. Schmidt-Radefeldt (Ed.), *Semiótica e lingüística portuguesa e românica. Homenagem a José Gonçalo Herculano de Carvalho* (pp. 93–98). Tübingen: Narr.
- Corominas, J., & Pascual, J. A. (1980–1991). *Diccionario crítico español e hispánico*. Madrid: Gredos.
- Darmesteter, A. (1967). *Traité de la formation des mots composés dans la langue française comparée aux autres langues romanes et au latin. Deuxième édition, vue, corrigée, et en partie refondue, avec une préface par Gaston Paris (1894)*. Paris: Librairie Honoré Champion.
- De Dardel, R., & Zamboni, A. (1999). L’interfixe -i- dans les composés proto-romans. Une hypothèse de travail. *Revue de linguistique romane*, 63, 439–469.
- Debrunner, A. (1917). *Griechische Wortbildungslehre*. Heidelberg: Carl Winter.
- Dressler, W. U. (2006). Compound types. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound words* (pp. 23–44). Oxford: Oxford University Press.
- Fabb, N. (1998). Compounding. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphology* (pp. 66–83). Oxford: Blackwell.
- Ferrari-Bridgers, F. (2005). Italian [VN] compound nouns: A case for a syntactic approach to word formation. In T. Geerts, I. van Ginneken & H. Jacobs (Eds.), *Romance languages and linguistic theory 2003: Selected papers from Going Romance 2003* (pp. 63–79). Amsterdam: John Benjamins. doi:10.1075/cilt.270.04fer
- Fruyt, M. (1990). La formation des mots par agglutination en latin. *Bulletin de la Société de Linguistique de Paris*, 85, 173–209. doi:10.2143/BSL.85.1.2013435
- Fruyt, M. (2002). Constraints and productivity in Latin nominal compounding. *Transactions of the Philological Society*, 100(3), 259–287. doi:10.1111/1467-968X.00099
- Harris, J. (1991). The exponence of gender in Spanish. *Linguistic Inquiry*, 22, 27–62.
- Haspelmath, M., & Sims, A. D. (2010). *Understanding morphology* (2nd ed.). London: Hodder/Hachette.
- Hatcher, A. G. (1951). *Modern English word-formation and Neo-Latin. A study of the origins of English (French, Italian, German) copulative compounds*. Baltimore, MD: Johns Hopkins Press.
- Herrera, M. T. (Ed.) (1996). *Diccionario español de textos médicos antiguos*. Madrid: Arco/Libros.

- Katamba, F. (1993). *Morphology*. New York, NY: St. Martin's Press.
- Kiparsky, P. (2009). Verbal co-compounds and subcompounds in Greek. In C. Halpert, J. Hartman & D. Hill (Eds.), *Proceedings of the 2007 Workshop on Greek Syntax and Semantics at MIT* (pp. 187–195). Cambridge, MA: MIT Working Papers in Linguistics.
- Maiden, M. (2008). Effects of word formation processes in Italian. Reflections on Maria Grossman and Franz Rainer (Eds.), *La formazione delle parole in italiano*. Tübingen: Niemeyer. *Rivista di Linguistica*, 20(2), 375–400.
- Maiden, M., Swearingen, A., & O'Neill, P. (2009). Imperative morphology in diachronic. Evidence for the Romance languages. In M. Dufresne, F. Dupuis & E. Vocaj (Eds.), *Historical Linguistics 2007. Selected papers from the 18th International Conference on Historical Linguistics, Montreal, 6–11 August, 2007* (pp. 99–108). Amsterdam: John Benjamins.
- Mallinson, G. (1986). *Rumanian*. Dover, NH: Croom Helm.
- Mascaró, J. (1986). *Morfologia*. Barcelona: Biblioteca Universitària/Enciclopèdia Catalana.
- Miller, D. G. (1993). *Complex verb formation*. Amsterdam: John Benjamins.
doi:10.1075/cilt.95
- Moyna, M. I. (2011). *Compound words in Spanish: Theory and history*. Amsterdam: John Benjamins. doi:10.1075/cilt.316
- Munthe, A. (1889). Observations sur les composés espagnols du type *aliabierto*. In *Recueil de mémoires philologiques présenté à M. Gaston Paris par ses élèves suédois le 9 août à l'occasion de son cinquantième anniversaire* (pp. 31–56). Stockholm: Imprimerie centrale.
- Nicholas, N., & Brian, J. (2009). Verbal dvandvas in Modern Greek. In C. Halpert, J. Hartman & D. Hill (Eds.), *Proceedings of the 2007 Workshop on Greek Syntax and Semantics at MIT* (pp. 171–185). Cambridge, MA: MIT Working Papers in Linguistics.
- Olsen, S. (2001). Copulative compounds: A closer look at the interface between syntax and morphology. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 2000* (pp. 279–320). Dordrecht: Kluwer. doi:10.1007/978-94-017-3724-1_11
- Rainer, F. (1993). *Spanische Wortbildungslehre*. Tübingen: Max Niemeyer.
doi:10.1515/9783110956054
- Ralli, A. (2009). Modern Greek V V dvandva compounds: A linguistic innovation in the history of the Indo-European languages. *Word Structure*, 2, 48–68. doi:10.3366/E1750124509000294
- Scalise, S. (1992). Compounding in Italian. *Rivista di linguística*, 4, 175–199.
- Spence, N. Ch. (1980). The gender of French compounds. *Zeitschrift für Romanische Philologie*, 96, 68–91. doi:10.1515/zrph.1980.96.1-2.68
- Spencer, A. (1991). *Morphological theory. An introduction to word structure in generative grammar*. Oxford & Cambridge, MA: Blackwell.
- Torres Cacoullós, R., & Bauman, J. (2018). Allative to purposive grammaticalization: A quantitative story of Spanish *para*. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 195–221). Amsterdam: John Benjamins. (in this volume). doi:10.1075/ihll.16.9tor
- Val Álvaro, J. F. (1999). La composición. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (pp. 4757–4841). Madrid: Espasa.
- Wälchli, B. (2005). *Co-compounds and natural coordination*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199276219.001.0001
- Whitney, W. D. (1941 [1879]). *Sanskrit grammar: Including both the classical language, and the older dialects of Veda and Brahmana*. Cambridge, MA: Harvard University Press.

Corpus

- ADMYTE: *Archivo Digital de Manuscritos y Textos Españoles*. Madrid: Micronet & Ministerio de Educación y Ciencia, 1992.
- Alonso Pedraz, M. (1986). *Diccionario medieval español: Desde las Glosas emilianenses y silenses (s. X) hasta el siglo XV*. Salamanca: Universidad Pontificia de Salamanca.
- Covarrubias Orozco, S. de. (2001 [1613]). *Suplemento al Tesoro de la lengua española castellana*, G. Dopico & J. Lezra (Eds.). Madrid: Polifemo.
- Covarrubias Orozco, S. de. (2006 [1611]). *Tesoro de la lengua castellana o española*, I. Arellano & R. Zafra (Eds.). Madrid: Centro para la Edición de Clásicos Españoles/Real Academia Española & Frankfurt am Main: Iberoamericana/Vervuert.
- Davies, M. *Corpus del español*. Provo, UT: Brigham Young University. <<http://www.corpusdelespanol.org>>
- Kasten, L., & Cody, F. (2001). *Tentative dictionary of Medieval Spanish*. New York: Hispanic Seminary of Medieval Studies.
- Kasten, L., & Nitti, J. (2002). *Diccionario de la prosa castellana de Alfonso X el Sabio*. New York: Hispanic Seminary of Medieval Studies.
- Menéndez Pidal, R., Lapesa, R., Seco, M., & García, C. (2003). *Léxico hispánico primitivo (siglos VIII al XII)*. Madrid: Espasa Calpe.
- Moliner, M. (1998). *Diccionario de uso del español*. Madrid: Gredos.
- Nebrija, A. de. (1973 [1495]). *Vocabulario romance en latín. Transcripción crítica de la edición revisada por el autor (Sevilla, 1516)*, G. J. MacDonald (Ed.), Philadelphia, PA: Temple University Press.
- Real Academia Española. (1884). *Diccionario de la lengua castellana por la Real Academia Española*. Madrid: Imprenta de D. Gregorio Hernando.
- Real Academia Española. (1979 [1726–1739]). *Diccionario de autoridades*. Madrid: Gredos.
- Real Academia Española. *Corpus diacrónico del español (CORDE)*. Madrid: RAE. <<http://www.rae.es>>
- Real Academia Española. *Corpus de Referencia del Español Actual (CREA)*. Madrid: RAE. <<http://www.rae.es>>
- Real Academia Española. *Nuevo tesoro lexicográfico de la lengua española*. Madrid: RAE. <<http://www.rae.es>>
- Seco, M., Andrés, O., & Ramos, G. (1999). *Diccionario del español actual*. Madrid: Aguilar.

Intersubjectification and textual emphasis in the use of *definite article* + *proper name* in Spanish

Miguel Calderón Campos

Universidad de Granada

This paper analyses the uses given to the structure [definite article + proper name] (e.g., *el Juan* ‘lit. the Juan’) in the history of Spanish (Calderón Campos, 2015). It employs the notions of “expressive value” of the definite article (Epstein, 1994, 2001) and of “intersubjectification” (Traugott & Dasher, 2002). At present this structure has a restricted colloquial use, with varying degrees of social perception and different pragmatic values depending on the communicative situation. However, in the past it was employed in learned texts with two possible values: (i) a discourse-pragmatic use with an anaphoric, non-judgemental value, equivalent to [‘the already mentioned’ + proper name]; and also a recognitional use, similar to the Latin *ille Socrates* ‘the well-known Socrates’.¹

Keywords: definite article; proper name; recognitional; intersubjectification; deixis; emphasis; Spanish

1. Introduction

The goal of the present paper is to analyse the distinct value of the structure [definite article + proper name], as shown in (1)–(6). Examples (1)–(5) have been extracted from 16th to 18th century written sources of learned texts. The first three examples appear in legal texts: more precisely, example (1) appears in a formal request to the Royal Chancery of Granada; (2) in the questioning

1. The present study is funded through the FFI2013-46207-P (MINECO) and FFI2017-83400-P (MINECO/AEI/FEDER,UE) research projects.

of a trial for slander and (3) in the questioning of a trial for injuries. Examples (4) and (5) belong to historical chronicles. Example (6), on the contrary, is an extract from contemporary spoken speech, taken from a conversation held in the city of Granada. In it, a father addresses his older son to instruct him to stand by the younger one, who is around two years old. Although all the cases present the same structure namely, [definite article + proper name], their use and pragmatic value is different. The goal of this paper is to describe the different uses of this structure: it will be shown that there are in fact three distinct uses of this structure.

- (1) [...] *dando mi parte en arrendamiento al Manuel Pérez una rehuerta [...] vino mi parte a tener la desgracia de ser el Pérez ahijado del citado escribano, con cuyo auxilio ha perdido dicha huerta y olivar [...] instando por que se hiciese otro nuevo reconocimiento para que el Pérez pagase los daños que de nuevo había causado [...]*
 '[...] and renting my part of a vegetable garden to [definite article +] Manuel Pérez [...] I had the misfortune that [definite article +] Pérez was the godson of the aforementioned notary, with whose help he lost the garden and olive grove [...] and I request that a new reckoning is made, so that [definite article +] Pérez pays for the damage he has caused [...]'
 (Provisión, 1782, 1r-1v)
- (2) *2ª. Y si saben que la motora del disturbio y quimera suscitada entre las referidas María Martínez y Agustina de Aranda [...] fue la Agustina de Aranda, la que provocó a la María Martínez con varias expresiones, ultrajándola y vilipendiándola, por ser la susodicha una mujer de genio intrépido, colérico, acostumbrada a tratar mal a todas las vecinas, por lo que está aborrecida en aquel pueblo.*
 '2nd. And whether they know that the one who caused the trouble and arose the dispute between the aforementioned María Martínez and Agustina de Aranda [...] was [definite article +] Agustina de Aranda, who provoked [definite article +] María Martínez with her various words, offending and disrespecting her, being the aforementioned a woman of wild and choleric temper, who is used to mistreat all her neighbours, which is why she is abhorred by the whole village'
 (Probanza sobre injurias, 1785, 13r)
- (3) [...] *luego que el Caminos salió de la prisión [...] lo estuvo asechando el don Bartolomé Sanches de noche [...] prevenido de armas ofensivas, como era una espada ancha o sable.*
 '[...] after [definite article +] Caminos was released from prison [...] he was stalked by [definite article +] don Bartolomé Sanches at night [...] armed with weapons, like a broad sword or a sabre'
 (Probanza contra B. Santiesteban, 1755, 6r)

- (4) [...] *e hirieron a Esteban Martín, lengua, e a otro cristiano [...] y si el Esteban Martín no se supiera curar, también muriera.*
 ‘[...] and so they injured Esteban Martín, interpreter, and another man [...] and if [definite article +] Esteban Martín had not had the skill to heal himself, he would have died too’ (G. Fernández de Oviedo, 1535–1557, *Historia general y natural de Indias*. CORDE)
- (5) [...] *que han osado enprender lo que nunca ni el Alixandre emprendió*
 ‘[...] who have dared to undertake what [definite article +] Alexander himself never undertook’
 (G. García de Santa María, 1499, *Traducción de la Crónica de Aragón*. CORDE)
- (6) *Ponte allí donde está el Ale*
 ‘Go stand there where [definite article +] Ale is’

In (1)–(4) articles have a discourse-pragmatic function. They precede the proper nouns, which have been previously mentioned in the text, and whose referent has already been introduced. They operate as a cohesive mechanism, which helps articulate the discourse at the textual level. The article highlights the presence of a proper name and reinforces the continuous presence of the referent throughout the discourse. All along, the identification of the referent is guaranteed by the unequivocal proper name.

Example (5), on the other hand, exhibits the recognitional use of the article, which precedes the name of a famous character of antiquity. *El Alixandre* is equivalent to “the famous Alexander, the Alexander known by all”. As is the case for examples (1)–(4), (5) is also an instance of a learned written style but, unlike in the previous examples, the use of the article does not require the previous mention of the referent. The article does not operate at the textual level any longer, nor does it refer to previously mentioned elements. Instead, it evokes a shared cultural universe known to both the author and the readers.

Lastly, (6) is a prototypical example of colloquial oral speech, reproducing a fragment of a conversation between a father and his sons. The function of the article in this case is to mark the noun phrase as informal, which is consistent with the communicative situation described. The article here is, therefore, a marker for informality or colloquiality. Of the uses described above, this is the only one that survives in currently spoken Spanish – subject to dialectal variation.²

This article is concerned with the analysis of these three uses of the definite article: (i) the informal use, (ii) the discourse-pragmatic use, and (iii) the recognitional use, from both a theoretical and an empirical viewpoint. On the theoretical

2. Apart from these, there also exists a formal journalistic use in which the article precedes the surname of famous female singers or writers: *la Caballé, la Pardo Bazán* (RAE, 2005, s.v. *el*).

side, I will try to discern whether the notions of “expressive value” of the definite article (Epstein, 1994, 2001) and “intersubjectification” (Carlier & De Mulder, 2010; Traugott & Dasher, 2002), which have been proposed to explain the origin of the definite article in the Latin demonstrative *ille*, can be applied to account for the use of the [definite article + proper name] structure in Spanish.

According to Epstein there is, in addition to a referential use of the article, an expressive use, which is “speaker-oriented” and “motivated by the speaker’s desire to portray the referent as discourse prominent” (Epstein, 2001, p. 186). According to Carlier and De Mulder (2010), following Traugott and Dasher (2002) and Traugott (1995, 2003, 2010), intersubjectification played a fundamental role in the transition from demonstrative to definite article, i.e. “the active role of the speaker to orient and to guide the hearer in his interpretational tasks” (Carlier & De Mulder, 2010, p. 265). In their opinion, in the grammaticalisation process that the definite article underwent, the distal demonstrative’s power to “mobilize specific knowledge shared by speaker and hearer” (De Mulder & Carlier, 2011, p. 529) was of paramount importance, in order to retrieve the referent when it is not immediately available or in the textual context.

We will examine how the use of the definite article before a proper name in Spanish can be explained by both concepts: the definite article as a marker of textual prominence and as an interactional strategy employed by the speaker to evoke previous knowledge shared by the hearer, thereby guiding the interpretation of the content.

The empirical basis of this study comprises three main sources:

1. The Real Academia Española’s Diachronic Corpus of Spanish (CORDE), from which I have selected texts mentioned in Section 2, comprising historical chronicles;
2. Transcriptions of unpublished 18th century court documents from the archives of the Royal Chancery of Granada, which are referenced in the Corpus section. For these I have transcribed the list of questions that witnesses were asked (examples (2)–(3)), as well as their answers (31)–(32), and some requests addressed to courts (1). This addition was motivated by the relatively small sample of texts of this nature in CORDE. The documents in CORDE and those in the Granada archive corpus have the following in common: they provide accounts that are either historical or judicial in nature. We opted for these text types because narratives tend to have several characters, which are first introduced and then appear again as the narration progresses. A narrative is thus an appropriate textual type to illustrate the several instances of the use of the noun phrases under analysis here.

3. Lastly, the colloquial examples of [definite article + proper name] have been collected either by myself for the city of Granada or by Carranza Brito (2008), Bize (2010) and Maradona (2000) in Mexico City, Santiago de Chile and Buenos Aires, respectively.

2. From the Latin distal demonstrative *ille* to the definite article

In Spanish, the definite article originates from the distal demonstrative *ille*. Although the grammaticalisation process initiated in Late Latin (Putzu & Ramat, 2001; Selig, 1992), fully-fledged definite articles appeared in the emerging Romance languages from the 9th century onwards (De Mulder & Carlier, 2011). In a document from San Millán de la Cogolla (year 800) there are 22 examples of *ille* used as definite article (Lapesa, 2000, p. 373). The *Glosas Emilianenses* and the *Glosas Silenses* (10th–11th c.), among the first texts written in Hispano-Romance, use the definite article (*elo, ela, lo, la*) extensively (Lapesa, 2000, p. 375).

Lyons (1999, pp. 161, 332) has argued that the shift from demonstrative to definite article is initiated when the referent is immediately accessible in the situation (exophoric) or in the context (endophoric). Deictic information is redundant in these two contexts of use, and therefore semantic weakening or “bleaching” can occur. Carlier and De Mulder (2010, p. 246) illustrate the loss of deictic meaning of *ille* with example (7), which proceeds from the 7th-century *Chronicle of the Merovingian Times*:

- (7) *Lucca castrum dirigunt, atque funditus subvertunt, custodes illius castris capiunt*

‘They go to the fort of Loches, they raze it to the ground and take the guardians of that fort prisoner’

(Fredegarius, *Continuations* § 25, cited by Carlier & De Mulder, 2010, p. 246)

According to Epstein (1994, p. 286, 2001, p. 186), the definite article developed from expressive uses of the Vulgar Latin demonstrative *ille*, as in (8). In Latin, definite reference to the participants was normally marked by zero determination (e.g., *cum auctoritate papae*, lit. ‘with authority of Ø Pope’), but when a writer wanted to attract the reader’s attention to a protagonist or an important object, he would use explicit markers (e.g., *cum auctoritate papae illius* ‘with authority of that Pope’). In (8), the content of the nominal expression is sufficient to identify the pope, and *ille* is used purely for emphasis (Carlier & De Mulder, 2010, pp. 249, 261).

- (8) *Septimo igitur regressus anno a Roma, cum auctoritate papae illius, Toronu redire disponit*
 ‘After seven years he quitted Rome and prepared with the pope’s sanction (lit. ‘the sanction of that pope’) to return to Tours’ (Gregorius the Great, *Historia Francorum*, cited by Trager, 1932, p. 172; Epstein, 1994, p. 286)

In Late Latin the use of demonstratives and other referential markers, such as *ipse* ‘this same’ and *praedicto* ‘the aforementioned’ (e.g., *ipse rex* ‘this same king’, *praedicto regi Pippino* ‘the aforementioned King Pippin’), increased in order to express definiteness and anaphoric relationships (Carlier & De Mulder, 2010, p. 253), even when the communication was not at risk of referential ambiguity. This increased frequency weakened its deictic and anaphoric functions. Consequently, the use of the deictic function of *ille* to indicate remoteness reduced and this item started being used to refer to prominent entities. Finally, the article ended up referring to entities that do not require being located in space or in discourse (Company, 2008, p. 176; Ortiz Ciscomani, 2009, pp. 334–367).

In the Romance languages, articles started being used when the entity referred to was highly individuated and, in particular, for concrete singular nouns such as *mirror*, *hat*, *sword*, etc.; (Company, 1991, p. 179). From these nouns, which are prototypical for their class, articles spread hesitantly but progressively to nouns referring to less individuated entities (De Mulder & Carlier, 2011, p. 524).

3. *Definite article + proper name in present-day Spanish*

As seen in (6) (and (9a)), certain colloquial Spanish varieties allow articles to precede proper names. This phenomenon applies to given names (e.g., *el Juan*, *la María*), surnames (e.g., *el Pérez*, *la Pérez*), to the combination of both (e.g., *la Lucía Sánchez*, *el Juan Gómez*), and to hypocorisms (e.g., *la Reme*, *la Lupe*, *el Juanra*, *la Charo*). In all these cases, the noun may or may not present a preceding article: *Juan/el Juan*, *María/la María*.³

The debate on the presence of an article preceding a proper name has often revolved around two questions:

1. Why is an article added to proper names when these are already intrinsically determined?
2. Does the proper name become a common noun when preceded by an article?

3. The present paper will not discuss those cases in which the article precedes monickers (e.g., *el Negro*, *el Pájaro*) or aliases (e.g., *el Cordobés*, *la Chunga*) due to space reasons.

The first question is of fundamental importance in order to understand the workings of the structure [definite article + proper name] in Spanish. We already know that the article can function as a form of introduction and update of a common noun, which is estimated to be known and accessible to the reader or speaker (Company, 2009, p. 791). Therefore, referents in noun phrases are unique and perfectly identifiable entities in a given communicative situation, in which shared knowledge between the communication participants is presupposed (Company, 2009, p. 794; RAE/ASALE, 2009, pp. 1042–1045).

Similarly, in specific contexts “we use proper nouns as if they were absolutely unique” (Lyons, 1999, p. 21). At first glance it might appear anti-economical to build a noun phrase with an article and a proper name, as both contribute similar information regarding the definiteness of the entity. The structure [definite article + proper name] can therefore be regarded as an overspecified phrase (Company, 2009, pp. 794–795). Such phrases are restricted in their use, but can be pragmatically exploited as long as there are unmarked constructions in the same grammatical functional space (Company, 2009, p. 841). This is precisely the case for the structure [definite article + proper name]. It appears as an overspecified noun phrase that is distinct from its counterpart without an article, and is restricted to colloquial situations. It is also easy to exploit pragmatically, which usually links it to evaluations of the referent (either positive or negative) and closely related to a presumably shared previous knowledge.

Arias Álvarez (2012), following Biber (1988) and Koch and Oesterreicher (2007), re-elaborates the parameters used to determine the degree of informality in communicative written exchanges. On the one hand, informality increases with a higher degree of closeness between the sender and the receiver, with the emotional involvement of the sender, and with the degree of dialogicity of the communication. On the other hand, informality decreases when the cultural status of the sender increases, when the number of receivers is high, when the purpose of communication is purely informative, and when the theme is highly fixed. In addition to these parameters, formality is usually increased by distance and by time.^{4,5}

The structure [definite article + proper name] is always marked as informal, i.e. it appears in communicative situations featuring a high degree of familiarity between the participants and a certain degree of emotional involvement. These

4. In Biber (1988) the degree of formality in telephone conversations is higher than in face-to-face exchanges.

5. For this reason, e-mail communications, which are meant to arrive in the same day, are usually more informal than messages sent by postal mail.

exchanges usually take place on a face-to-face basis, with not strictly informative (or sometimes informative-emotional) objectives and little fixation on theme. In such situations the shared context is large and inferences derived from that context may play an important role.

This informality is constant and a defining feature of the construction under discussion. It is often the only differential feature in the structure: the only difference between (6), repeated below as (9a), and (9b) is the higher degree of informality that the article expresses, which is to be expected in a conversation between parents and children. In this case, the article does not carry any evaluative connotations nor does it help identify the referent, who is well known to the members of his family and visible to all in that particular situation:

- (9) a. *Ponte allí donde está el Ale*
 ‘Go stand there where [definite article +] Ale is’
 b. *Ponte allí donde está Ale*
 ‘Go stand there where Ø Ale is’

In other cases, the article preceding a proper name is linked to subjective interpretations of the referent, as in the following examples, which permit a laudatory (10), derogatory (11) and ironic (12) interpretation, respectively:

- (10) *A veces me hace enojar la Daisy, pero es muy noble, siempre está pendiente de mí*
 ‘Sometimes [definite article +] Daisy makes me mad, but she is kind-hearted, she is always looking after me’ (cited by Carranza Brito, 2008).
- (11) *¿Qué se cree el Pedro Ángel que no llega? ¿De veras cree que no tenemos nada mejor que hacer o que lo vamos a esperar toda la vida?*
 ‘Who does [definite article +] Pedro Ángel think he is to be so late? Does he really believe that we have nothing better to do, or that we will wait for him indefinitely?’ (cited by Carranza Brito, 2008)
- (12) – *Andrés: ¿Tú estás bien? ¡Y el Pablo es un grande!*
 – *Mariana: ¡Chistoso!*
 ‘– Andrés: Are you all right? And [definite article +] Pablo is such a great guy!
 – Mariana: You are so funny!’

Examples (10) and (11) were collected by Carranza Brito (2008) in spontaneous conversations that took place in Mexico City. There is no doubt about the positive subjective value of (10) and the negative subjective value of (11). Example (12), extracted from the Chilean film *La vida de los peces* (Bize, 2010), needs further contextual explanation. The conversation takes place between Andrés and Mariana, who have been friends since childhood. Andrés just arrived to Santiago de

Chile after fifteen years abroad. They reminisce about their lives and Andrés mentions Pablo, Mariana's husband, in apparently favourable terms: *¡Y el Pablo es un grande!* 'And [definite article +] Pablo is such a great guy!'. He actually knows that Pablo is not a great guy at all, and that he is probably the same rascal he always was. Later, Mariana will confirm this by stating that her husband does not care much for her, as he spends most of his time playing with his computer.

The evaluative interpretation in these cases is very sensitive to context, be it verbally explicit (e.g., *la Daisy [...] es muy noble* '[definite article +] Daisy [...] she is kind-hearted', *¿Qué se cree el Pedro Ángel que [...]?* *¿De veras cree que [...]?* 'Who does [definite article +] Pedro Ángel think that [...]? Does he really think that [...]?') or subsumed in the shared knowledge by the sender/speaker and receiver/hearer, and required for the interpretation to go in a certain way: as such, the intended ironic interpretation in (12) is closely linked to the shared knowledge that Andrés and Mariana have on Pablo.

The use of the definite article is not the sole cause of an evaluative interpretation, since it could also happen in its absence. Nevertheless, the article helps the addressee to realise that the speaker is making an expressive use of the structure, and suggesting a mobilisation of shared knowledge in order to generate a correct interpretation of the statement.

In my opinion, the definite article can thus perform two functions. On the one hand, it can impart the feature [+informal] to the statement that is being made. In cases like (9a) it is the only possible value for the article preceding the proper name. On the other, it can help the hearer to invoke certain knowledge that is shared with the speaker. The article adverts the addressee that a stylistic-expressive use of the construction might be taking place, namely a pragmatic exploitation of the overspecified phrase.

Consequently, the article restricts the use of the noun phrase to informal situations, but at the same time extends its possibilities for pragmatic and expressive exploitation, by warning the hearer that inferences will be needed for successful understanding. The use of the definite article as a warning for the activation of inferences is a clear manifestation of the intersubjective dimension, as pointed out by Carlier and De Mulder (2010): the speaker is aware of the difficulties that the hearer may have when interpreting a statement and provides help by giving out clues, which in this case is the use of a definite article before a proper name.

When dealing with the reasons behind the use of a definite article preceding a proper name, we must ask ourselves whether this article has any identificational value at all, and therefore whether it contributes to making an unequivocal denotation. Such identificational value has been advocated for by Gary-Prieur (1994, pp. 98–104) and Devís Márquez (2009, pp. 456, 479):

Whenever an article, in its singular form, appears alongside a given name or a surname (*la María, el Pedro, el Sánchez, la Callas*), since denotation is oriented toward an individual entity, the determiner acquires a certain identificational value. (Devis Márquez, 2009, p. 479)

However, in my opinion, the article does not perform an identificational function in these cases, as this task is already carried out by the proper name. In example (9a) the father does not employ the article to inform his older son of which specific Ale he is referring to. The only possible Ale is visible to all and absolutely unambiguous. The same can be said for the examples in (13)–(15):

- (13) – *El otro día vi a tu nuevo nieto. Es idéntico al mayor.*
 – *Sí, se parece mucho al Carlos.*
 ‘I saw your new grandchild the other day. He looks exactly like the older one.
 – Yes, he resembles [definite article +] Carlos very much’
- (14) [...] *la mamá de Juan, la Inés, era siempre la primera en la clase.*
 ‘[...] Juan’s mum, [definite article +] Inés, was always the best student of her class’
- (15) *Vieja, vamos a la cancha, vamos que hoy juega el Diego. Porque yo soy el Diego y soy de los que me llaman así [...] les había fallado a ellos, a los que habían llenado la Bombonera todos los domingos para ver a Maradona, a el Diego. Cuando terminó el partido, un tristísimo dos a dos [...] se pusieron a gritar: ¡El Diego no se va!, ¡el Diego no se va!*
 ‘Woman, let’s head for the field, come on, today [definite article +] Diego plays. Because I am [definite article +] Diego and I am one of those who call me that [...] I had failed them, the ones who crowded *la Bombonera* every Sunday to see Maradona, [definite article +] Diego. Once the game was over, a paltry two-two draw [...] they started yelling: [definite article +] Diego will stay!, [definite article +] Diego will stay!’
 (Yo soy el Diego, 2000. CREA)

I collected examples (13) and (14) in Granada, while example (15) proceeds from the autobiography of the Argentinian football player Diego Armando Maradona. In none of these cases the article is essential to identify the referent and is, as such, semantically unnecessary. In (13), there is no doubt that Carlos is the older grandchild and in (14), when Inés is mentioned, the hearer already knows who she is, since she has been described as Juan’s mother. Similarly, example (15) is clear-cut: the player calls himself *el Diego*, obviously not to ensure his identification, but to add an expressive value, which in this context can be interpreted as belonging to a social group, to a neighbourhood, and to its team: ‘I am [definite article +] Diego, and I belong to those who call me that,

not to those who call me Maradona.’ This example reveals that the article activates context-dependent, subjective interpretations, whereas the identificational function falls on the anthroponym.

According to Longobardi (1994, pp. 646–652), the article has an expletive value in Italian examples like *il Gianni mi ha telefonato* ‘[definite article +] Gianni phoned me’, and it cannot be argued that the proper name becomes a common noun.⁶ On the contrary, when proper names are modified by a restrictive relative clause (*il Gianni che conoscevo non esiste più* ‘The Gianni I knew no longer exists’) they become common nouns. In these cases, and in similar ones like *esta no es la María que yo conozco* ‘this is not the María that I know’ (Roca Urgell, 1996, p. 137), the proper name, as common nouns do, presupposes a class which already comprises other *Giannis* and *Marías*. The same can be said when the proper name is joined by a specifying adjective, considering that then “the reference is limited to an aspect, a perspective, or a quality” (Lapesa, 2000, p. 437), as in *el César estratega* ‘the strategist Caesar’. In these cases, proper names lose their unitary referencing quality and “come close, in a certain way, to the common noun” (Lapesa, 2000, p. 437). It is as if the individuality of Caesar is segmented in the several aspects of his character (Brucart, 1999, pp. 409–411; Devís Márquez, 2009, p. 450). These uses, in which the proper name is joined by restrictive complements, are termed by Fernández Leborans (1999, p. 103) “non-prototypical” and are different from the typical uses of the referential (*Juan está estudiando* ‘Juan is studying’), vocative (*¡ven aquí, Juan!* ‘come here, Juan!’) and denominative anthroponyms (*me llamo Juan* ‘I’m called Juan’).

In summary, in present-day Spanish, in the varieties discussed here, the presence of a determiner endows the noun phrase [definite article + proper name] with the [+informal] marker, and prepares the addressee for the activation of inferences from the shared context (intersubjectification). In these cases like (6) (repeated as (9a)) the determiner has an expletive value, is not necessary to identify the referent, and does not cause the proper name to become a common noun.

4. The definite article as a marker of discourse prominence

This section will analyse examples (1)–(4), repeated below as (22)–(25), in which the structure [definite article + proper name] appears in narrative passages from legal (22)–(24) and historical texts (25). The question should be raised why this

6. This opinion is shared by Fernández Leborans (1999, p. 116), Laca (1999, p. 924), Leonetti (1999, p. 880), and Rigau (1999, pp. 320–321).

overspecified structure is employed so frequently in legal and historical texts. Does the article contribute to identify the referent or is this function performed by the proper name? If the article is not an identifier, what are its textual effects? Can this structure be used, as in present-day Spanish, to express a positive or negative evaluation of the referent? I intend to show that in examples (22)–(25), unlike in the examples of Section 3 (which were collected from spontaneous present-day spoken Spanish), the use of the article is not informal nor does it carry an evaluative description of the referent. On the contrary, the article helps the reader to understand that the referent is important in that fragment of text, and that it will be present along the flux of the narration (Company, 2009, p. 858). This is, therefore, a discourse-pragmatic use, of an emphatic nature, and aimed at pointing out its referential persistence. This mechanism is restricted to the text, and it reinforces the continuity of the referent in the discourse, as it is already identifiable without the need for an article.

Before examining the pragmatic value of the article of these cases, I believe some preliminary observations are in order about the moment in the discourse at which it becomes acceptable for the article to precede a proper name. As mentioned previously, in colloquial uses of this structure, the article may appear at the first mention of a person, with no previous introduction required, as in (9a) and (12). In examples (1)–(4)/(22)–(25), on the contrary, the article can only appear after the person has been introduced. The preceding context to these examples, in which the characters are mentioned for the first time, is given in (16)–(19), which do not exhibit the structure under discussion.

- (16) [...] *me querello de Manuel Pérez de Alcázar, el menor, y de Francisco Molina, ambos vecinos de Loxa y este, escribano de ella, y digo que [...]*
 ‘[...] I bring action against **Manuel Pérez de Alcázar**, the younger, and against **Francisco Molina**, both neighbours of the town of Loxa, and being this one the scribe of the town, and I say that [...]
 (Provisión, 1782, 1r)
- (17) *Por las preguntas siguientes se examinarán los testigos que se presenten por parte de Juan Pérez, vezino de la villa del Marchal, marido de María Martínez, en los autos sobre ynjurias que sigue con Miguel de Moya, marido de Agustina de Aranda*
 ‘The following questions will serve to examine the witnesses brought by **Juan Pérez**, neighbour of the village of **Marchal**, husband of **María Martínez**, in the slander case which implicates **Miguel de Moya**, husband of **Agustina de Aranda**’
 (Probanza sobre injurias, 1785, 13r)
- (18) *Por las preguntas siguientes se examinarán los testigos que se presentaren por parte de Antonio Caminos, vecino de la ciudad de Vélez, en el pleito [...] que sigue contra don Bartolomé Santiesteban Sánchez, vecino de dicha ciudad*

‘The following questions will serve to examine the witnesses brought by **Antonio Caminos**, neighbour of the town of Vélez, in the case against **don Bartolomé Santiesteban Sánchez**, neighbour of said town’

(*Probanza contra B. Santiesteban*, 1755, 6r)

(19) *Los indios [...] se pusieron en armas e hirieron a Esteban Martín, lengua*

‘The Indians took up arms and injured Esteban Martín, the interpreter’

(G. Fernández de Oviedo, 1535–1557, *Historia general y natural de Indias*. CORDE)

Notwithstanding the previous examples, observe that in some historical chronicles, as illustrated in (20), an indefinite article preceding a proper name can be used instead in the introduction (*un Francisco Mercader*) and later the definite one in subsequent mentions (*al Francisco Mercader*):

(20) *Fue el que levantó aquella ciudad un Francisco Mercader, pobre escudero, y como lo supo D. Luis Hurtado de Mendoza... fue luego con mucha gente de armas a Baza y tomola y allanáola y descuartizó al Francisco Mercader*

‘The one who built that town was [indefinite article +] **Francisco Mercader**, a poor squire, and after D. Luis Hurtado de Mendoza heard about it [...] he went to Baza with many armed men and took it [the town], laid it to waste and quartered [definite article +] **Francisco Mercader**’

(A. de Santa Cruz, ca. 1550, *Crónica del emperador Carlos V*. CORDE)

It should further also be pointed out that for the first mention of a character, no examples like (21) have been encountered, which I therefore consider to be impossible, given that, the presence of the definite article in the introduction would impart to (21) a degree of colloquiality incompatible with the formal tone expected from a legal request.

(21) **Me querello del Manuel Pérez, el menor, vecino de Loxa [...]*

‘I bring action against [definite article +] **Manuel Pérez** the younger, neighbour of Loxa [...]

In the following, I will show that in examples (1)–(4), repeated below as (22)–(25), the [definite article + proper name] structure does not necessarily imply admiration nor contempt for the person mentioned. As such, I do not share the opinion of Reynoso Noverón (2008) and Ortiz Ciscomani and Reynoso Noverón (2012), who defend that the use of a definite article before an anthroponym intends to provide a negative view of the people mentioned (Reynoso Noverón, 2008, p. 2125), nor that its use is linked to “contexts of noticeable social degradation or intense negativity” (Ortiz Ciscomani & Reynoso Noverón, 2012, p. 2318).

- (22) [...] dando mi parte en arrendamiento **al Manuel Pérez** una rehuerta [...] vino mi parte a tener la desgracia de ser **el Pérez** ahijado del citado escribano, con cuyo ausilio ha perdido dicha huerta y olivar [...] instando por que se hiciese otro nuevo reconocimiento para que **el Pérez** pagase los daños que de nuevo había causado [...]
 '[...] and renting my part of a vegetable garden to [**definite article** +] **Manuel Pérez** [...] I had the misfortune that [**definite article** +] **Pérez** was the godson of the aforementioned notary, with whose help he lost the garden and olive grove [...] and I request that a new reckoning is made, so that [**definite article** +] **Pérez** pays for the damage he has caused [...]'
 (Provisión, 1782, 1r-1v)
- (23) 2ª. Y si saben que la motora del disturbio y quimera suscitada entre las referidas **María Martínez** y **Agustina de Aranda** [...] fue **la Agustina de Aranda**, la que provocó a **la María Martínez** con varias expresiones, ultrajándola y vilipendiándola, por ser la susodicha una mujer de genio intrépido, colérico, acostumbrada a tratar mal a todas las vecinas, por lo que está aborrecida en aquel pueblo.
 '2nd. And whether they know that the one who caused the trouble and arose the dispute between the aforementioned **María Martínez** and **Agustina de Aranda** [...] was [**definite article** +] **Agustina de Aranda**, who provoked [**definite article** +] **María Martínez** with her various words, offending and disrespecting her, being the aforementioned a woman of wild and choleric temper, who is used to mistreat all her neighbours, which is why she is abhorred by the whole village'
 (Probanza sobre injurias, 1785, 13r)
- (24) [...] luego que **el Caminos** salió de la prisión [...] lo estuvo asechando **el don Bartolomé Sanches** de noche [...] prevenido de armas ofensivas, como era una espada ancha o sable.
 '[...] after [**definite article** +] **Caminos** was released from prison [...] he was stalked by [**definite article** +] **don Bartolomé Sanches** at night [...] armed with weapons, like a broad sword or a sabre'
 (Probanza contra B. Santiesteban, 1755, 6r)
- (25) [...] e hirieron a **Esteban Martín**, lengua, e a otro cristiano [...] y si **el Esteban Martín** no se supiera curar, también muriera.
 '[...] and so they injured **Esteban Martín**, interpreter, and another man [...] and if [**definite article** +] **Esteban Martín** had not had the skill to heal himself, he would have died too'
 (G. Fernández de Oviedo, 1535–1557, *Historia general y natural de Indias*. CORDE)

Although the negative interpretation may be acceptable in some contexts, it does not appear to be a constant, as is exemplified by (26) and (27), the former of which has been cited by Ortiz Ciscomani and Reynoso Noverón (2012):

- (26) [...] *juntaronse unos tres principales jndios con el Alonso Ortiz [...] y el Alonso Ortiz es conocido quien es*
 ‘[...] three leading Indians met with [definite article +] Alonso Ortiz. [...] and [definite article +] Alonso Ortiz is well known’
 (DLNE, 1555, pp. 138–139)
- (27) *Y de enojo que dél tenjan comenzaron a pjntar contra el bueno de Diego Ramírez, levantandole que havja tomado tantos mill pesos a esos jndios [...]. Y venjendo esto a notisia del Diego Ramírez, escrjbiome*
 ‘And out of anger towards him, they started telling tales of good Diego Ramírez, accusing him of taking a thousand pesos from those Indians [...]. And when [definite article +] Diego Ramírez found out about this, he wrote to me’
 (DNLE, 1555, p. 138)

(26) is taken from a letter written by a friar in defence of a royally-appointed inspector, Diego Ramírez, and against overseer, Alonso Ortiz de Zúñiga, who had conspired with some Indians to falsely accuse Ramírez. In his letter, the friar proves the innocence of Ramírez. In this particular context it could indeed be interpreted, as has been suggested by Ortiz Ciscomani and Reynoso Noverón (2012, p. 2321), that the author of the letter uses the article in order to express his negative opinion of Alonso Ortiz. In the same letter, however, Diego Ramírez is also mentioned with an article, as demonstrated in (27). This example proves unequivocally that the presence of an article before a proper name does not necessarily imply a negative opinion, since this structure can also be employed to refer to highly regarded characters such as “good Diego Ramírez”.

Indeed, when using a large enough corpus, it becomes apparent that the structure with a definite article does not necessarily carry derogatory overtones.⁷ García Gallarín (1983, pp. 58–59), for instance, analysed 41 occurrences of this construction from the 16th and 17th centuries: in 23 cases of these, the structure under study appeared in contexts revealing a favourable attitude towards the person mentioned, while only in 12 cases the author criticised the behaviour of the subject. The rest of the cases were not easy to interpret as either positive or negative.

Similar conclusions can be reached for example (23), in which the article precedes both the names of Agustina de Aranda, who is said to be bad-tempered and known to be abusive, and of María Martínez, the woman who suffered Agustina’s attack. In the first case, one might interpret the use of the article as derogatory, but there is no doubt that in María Martínez’ case the use of the article carries no evaluative weight.

7. Ortiz Ciscomani and Reynoso Noverón (2012, p. 2317) only found eight examples in their corpus.

The same can be also said of example (24), where both the aggressor, don Bartolomé Sánchez, and the victim, Antonio Caminos, have an article preceding their names. In (25), on the other hand, the author does not provide a positive nor negative evaluation of Esteban Martín's behaviour, but instead simply narrates how he was injured and then able to heal himself.

The instances in (28) and (29) are, respectively, a case of negative (*el Montalvo, el Pedro Villafuerte*) and positive evaluation (*el Pedro de Alvarado, el Francisco Montero*):

- (28) [...] *estando el gobernador echado en su cama, e teniendo la guardia de su persona aquel Montalvo, que era uno de la conjuración, dio entrada al Pedro Villafuerte, quedando a la puerta del gobernador el Montalvo y el Teniente Porras, haciendo espalda y favor al principal traidor. El cual [...] le dio cinco puñaladas muy presto [...] y como el Villafuerte entendi6 que era sentido, se descabuy6 [...]*

'[...] while the governor laid in his bed, and Montalvo was his guard, who was one of the conspirators, he let [definite article +] Pedro Villafuerte in, while [definite article +] Montalvo and lieutenant Porras remained by the door, covering him and easing the task for the main traitor. Who [...] stabbed him five times very quickly [...] and, as [definite article +] Villafuerte thought he had been heard, he ran away [...]

(G. Fernández de Oviedo, 1535–1557, *Historia general y natural de las Indias*. CORDE)

- (29) [...] *sepan que el Pedro de Alvarado fue un hidalgo muy valeroso [...] e asimismo el Francisco de Montejo, hidalgo de mucho valor, que fue gobernador y adelantado de Yucatán.*

'[...] let it be known that [definite article +] Pedro de Alvarado was a brave nobleman. [...] and also [definite article +] Francisco de Montejo, a nobleman of great value, who was governor and judge in Yucatan'

(B. Díaz del Castillo, ca. 1568–1575, *Historia verdadera de la conquista de la Nueva España*. CORDE)

Examples (23)–(29) corroborate that the presence of the article does not imply a subjective evaluation in these historical narrative and legal texts. In view of this, we must return to the question what the function of the determiner might have been. In an attempt to answer this, I will provide the wider context for example (24) in (30), which contains the full question that was asked to the witnesses in this trial at El Marchal (Granada):

- (30) *Y si saben que el referido don Bartolomé Santisteban Sánchez, llebado de el odio y mala voluntad que contra el referido Antonio Caminos tenía, nacido de la querella que contra el susodicho hauía dado por haver cogido limones en la huerta de el don Bartholomé, para lo que dice no dio su licencia, profirió*

en algunas ocasiones que se la había de pagar. Y que en consecuencia de estas amenazas, luego que el Caminos salió de la prisión, lo estuvo asechando el don Bartolomé Sanches de noche, preuenido de armas ofensivas, como una espada ancha o sable, lo que repitió en algunas noches hasta que el día doce, yendo el Caminos a recogerse a sus casas, salió el don Bartholomé de el saguán de sus casas y llamó diferentes veces a el Caminos hasta que, haviéndole respondido, le dijo el don Bartholomé que cómo era hombre tan ruin, con otras injurias semejantes. Y haviéndole asido de el cuello de la ropa, le tiró un golpe a dicho Caminos con dicha espada o sable y le hirió grauemente en el brazo [...]

‘And whether they know if the aforementioned don Bartolomé Santisteban Sánchez, full of hate and bad intentions towards the aforementioned Antonio Caminos, born from the dispute that against him was filed for stealing lemons from [definite article +] don Bartholomé’s orchard, for which, he stated, he had not given permission, declared on several occasions that he should be compensated for it. And that, following these threats, after [definite article +] Caminos was released from prison, he was stalked by [definite article +] don Bartolomé Sanches at night, armed with weapons, such as a broad sword or a sabre, which he did on several nights until on the twelfth day, when [definite article +] Caminos was on his way home, [definite article +] don Bartholomé came out of his house and called several times out to [definite article +] Caminos. Having responded to him, [definite article +] don Bartholomé told him that he was contemptible, and other similar insults. And having grabbed him by his collar, he hit the aforementioned Caminos with said sword or sabre, and hurt badly his arm [...]

(Probanza contra B. Santiesteban, 1755, 6r)

If the text in (30) is read without any articles, it becomes apparent that the determiner plays no identification role since the unequivocal reference to don Bartolomé and Antonio Caminos is sufficiently performed by the anthroponyms. It has no evaluative function either, as the article precedes both the aggressor and his victim, Antonio Caminos. Instead, the writers of these questions for the witnesses place the article before the name of the two main characters under trial for emphasis or prominence, and to indicate that they will be used again throughout the narration. It is demanded from the witnesses that their attention be focused on these characters, and that they remember everything they can about them. At no point is there any doubt on the ability of the witnesses to identify don Bartolomé or Caminos. Instead, an overspecified structure is used to emphasise the relevance of these characters at a particular point in the discourse. To use a typographical analogy, the author is underlining the name of the main characters in the action. The article has thus an in-text value here, in order to highlight significant entities. It does not help identify the main characters, but instead makes them prominent at this point in the discourse.

This expressive value of the definite article as marker of discourse prominence arises, according to Epstein (1993, p. 129) from its demonstrative origins, since the demonstrative is the linguistic equivalent of the act of pointing (Carlier & De Mulder, 2010, p. 248). Keniston (1937, p. 226) states that, in cases such as these, the article is equivalent to *el dicho* ‘the aforementioned’ or the demonstrative *este* ‘this’. In sum, the authors of legal texts and of historical chronicles had three resources available to emphasise a given character in the text: (i) *el dicho*, (ii) the definite article, and (iii) the demonstrative *este*.⁸ In our corpus, there are frequent co-occurrences of *el dicho*⁹ and [definite article + proper name], as seen in (30) and in (31):

- (31) [...] *el testigo vido que el Francisco Gonzales estaba sobre una tapia [...] y que desde ella le decía al dicho D. Antonio era un marrano [...] y encaminándose el citado D. Antonio hacia la tapia, con una caña o palo en la mano, arrancó el dicho Francisco González una piedra de la tapia para tirársela al D. Antonio*
 ‘[...] the witness saw that [definite article +] Francisco Gonzales was on a wall [...] and from the wall he said to the aforementioned D. Antonio that he was a *marrano* (secret jew) [...] and after that the aforementioned D. Antonio went to the wall with a stick or cane in his hand, and the aforementioned Francisco González took a stone from the wall to throw it at [definite article +] D. Antonio’
 (*Querella criminal. Fiñana, 1746, 3v-4r*)

The corpus used for this study also offers several examples of the demonstrative *este* with a non-restrictive value (Lyons, 1999, p. 122), as in (32) in which the demonstrative does not help to distinguish this Rueda from other Rueda’s being mentioned, but instead it is a way to emphasise that particular character in that fragment of discourse.

- (32) *el juez [...] llamó a Francisco de Rueda, testigo [...] y en efecto, así es que hallándose este Rueda en la ocasión del lance en su casa [...] no oyó más palabras*
 ‘the judge [...] summoned Francisco de Rueda as a witness [...] and as [demonstrative +] Rueda was present in the fight that took place in his house [...] he heard no other testimony’
 (*Probanza contra A. Zambrana, 1779, 14v*)

-
8. The demonstrative *aquel* is also used, as in example (28): *aquel Montalvo*.
 9. This also includes its synonyms *el citado*, *el susodicho*, *el referido*.

Present-day Spanish admits *este* and especially *el dicho* (and its synonyms) as mechanisms to achieve text emphasis, and rejects the use of the definite article for this purpose. For contemporary Spanish speakers, examples like *le dijo el don Bartolomé* ‘[definite article +] don Bartolomé told him’ or *llamó a el Caminos* ‘he called out at [definite article +] Caminos’ in (30) have clear informal overtones that render them inappropriate for the cultured prose. In other words, it is the informal character of the noun phrase [definite article + proper name] that has prevailed over the textual emphasis function.

In conclusion, the definite article was used in the written language as a marker of discourse prominence (Epstein, 2001) for characters who were important in a particular sequence of narration. Crucially, the use of the determiner did not imply a positive or negative evaluation of the character mentioned.

5. The recognitional use of *definite article + proper name*

I call “recognitional use” the value of the definite article in example (5), repeated below as (33). In this example, the determiner is used before a famous character of Antiquity, whom the writer presumes to be known to readers. The people mentioned include very well-known biblical characters e.g., *el Gabriel* in (34), heroes as in (33), sages from ancient times (35), or scholars in some field of knowledge (36):

- (33) [...] *que han osado enprender lo que nunca ni el Alixandre emprendió*
 ‘[...] who have dared to undertake what [definite article +] Alexander
 himself never undertook’ (G. García de Santa María, 1499, *Traducción de la*
Crónica de Aragón. CORDE)
- (34) *Fue tu quartä alegría*
quando te dixo, María,
el Grabiél
 ‘It was your fourth joy.
 When you were called forth, Mary,
 by [definite article +] Gabriel’ (Libro de Buen Amor, estrofa 38)
- (35) [...] *y el Aristotel dize qu’este mundo es pequeño [...] diziendo qu’el*
Aristóteles pudo saver muchos secretos d’estos del mundo a cabsa de
Alixandre Magno
 ‘[...] and so [definite article +] Aristotle says that this is a small world [...] saying the [definite article +] Aristotle might have come to know many secrets of the world because of Alexander the Great’
 (1498, *Relación del tercer viaje de Colón*. CORDE).

- (36) *Pues no presume de necio
aquel Francisco Aretino
ni aun el Ángelo de Arecio,
su natural y vecino*

'He does not boast being a dunce
[demonstrative +] Francisco Aretino
nor does [definite article +] Ángelo de Arecio,
his fellow countryman and neighbour'

(F. de Ávila, 1508, *La vida y muerte o Vergel de discretos*. CORDE)

The recognitional use of *el* is similar to its informal use to the extent that it does not require the referent to be mentioned previously. Nevertheless there is one significant difference: in the informal use, the value of the noun phrase is interpreted using specific knowledge shared by speaker and hearer, whereas in the recognitional use, the interpretation is based on stereotypical cultural knowledge shared by a wider community.

This recognitional use was already present in Latin, where the demonstrative *ille* followed by a proper name meant 'the illustrious, the well-known', as in *Socrates ille* 'famous Socrates' or *ille Achilles* 'famous Achilles' (Glare, 1982, s.v. *ille*). De Mulder and Carlier (2011, p. 242) interpret this use as a manifestation of the subjective value that *ille* had in Latin. In this specific case, the demonstrative is used to express respect toward the referent.

The Latin-flavoured demonstrative *aquel* is already present in *La Celestina* (37) and was maintained as a resource of learned written style for a long time, as illustrated by (38) from the 19th century. Notice moreover that in example (36), the 16th-century writer used both the demonstrative *aquel* (*aquel Francisco Aretino*) and the definite article *el* (*el Ángelo de Arecio*) for the recognitional function.

- (37) *Pues menos podrás decir, mundo lleno de males, que fuimos semejantes en
pérdida aquel Anaxágoras y yo [...]*

'Less canst thou say, thou world replenished with evil, that [demonstrative
+] Anaxagoras and I were alike in our loss [...]

(F. de Rojas, ca. 1499–1502, *La Celestina*. CORDE)

- (38) *¡Ah! El amor fue el talón vulnerable de aquel Aquiles*

'Alas! Love was the vulnerable heel of [demonstrative +] Achilles'

(P. A. Alarcón, 1874, *La Alpujarra*. CORDE)

6. Conclusions

The analysis put forward in this paper has focused on three uses of [definite article + proper name]. Out of these, only one persists in certain varieties of present-day

Spanish: the informal use in which the article may help activate inferences in the addressee's mind, which are required to correctly interpret the evaluation that the speaker makes about the referent. The other two uses, in contrast, have become obsolete: in the first one, the article helps to highlight the textual prominence of relevant characters in a particular narrative sequence; while the other has a *recognitional* function, as the determiner tends to precede illustrious characters of the classical world.

From a theoretical perspective, the inferential effect of the article has been shown to be linked to the intersubjective dimension that Carlier and De Mulder (2010) consider to be a feature of the grammaticalisation process experienced by the Latin demonstrative *ille*.

The highlighting effect of the article has already been observed by Epstein (2001, p. 186) as characteristic of Latin *ille* in text sequences in which the demonstrative progressively loses its original deictic value. Several features of the demonstrative *ille* in Late Latin, present at the time when it started to weaken semantically, survived thus in the Spanish definite article. In Latin, the new pragmatic values of *ille* arose in overspecified phrases (*illius castris* (7), *cum auctoritate papae illius* (8)). Similarly, in Spanish, the overspecification of the combination [definite article + proper name] triggers the activation of pragmatic values of the noun phrase.

References

- Arias Álvarez, B. (2012). Construcción del discurso femenino novohispano: Inmediatez y distancia comunicativa. In E. Montero Cartelle & C. Manzano Rovira (Eds.), *Actas del VIII Congreso Internacional de Historia de la Lengua Española* (Vol. 2; pp. 1195–2022). Santiago de Compostela: Meubook.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511621024
- Brucart, J. M. (1999). La estructura del sintagma nominal: las oraciones de relativo. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 1; pp. 395–522). Madrid: Espasa.
- Calderón Campos, M. (2015). El antropónimo precedido de artículo en la historia del español. *Hispania*, 98(1), 79–93. doi:10.1353/hpn.2015.0003
- Carlier, A., & De Mulder, W. (2010). The emergence of the definite article: *ille* in competition with *ipse* in Late Latin. In K. Davidse, L. Vandelanotte & H. Cuyckens (Eds.), *Subjectification, intersubjectification and grammaticalization* (pp. 241–275). Berlin: De Gruyter Mouton. doi:10.1515/9783110226102.3.241
- Carranza Brito, M. R. (2008). Artículo ante nombre propio: Matices de significado. *LL Journal*, 3(2). Retrieved from <<http://ojs.gc.cuny.edu/index.php/lljournal/article/view/418/425>>
- Company Company, C. (1991). La extensión del artículo en el español medieval. *Romance Philology*, 44(4), 402–424.

- Company Company, C. (2008). La creación y generalización del artículo en español. *Manual de gramática histórica* (pp. 175–186). Ciudad de México: UNAM.
- Company Company, C. (2009). Artículo + posesivo + sustantivo y estructuras afines. In C. Company Company (Ed.), *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal* (pp. 761–880). Ciudad de México: Fondo de Cultura Económica/UNAM.
- De Mulder, W., & Carlier, A. (2011). The grammaticalization of definite articles. In H. Narrog & B. Heine (Eds.), *The Oxford handbook of grammaticalization* (pp. 522–535). Oxford: Oxford University Press.
- Devis Márquez, P. P. (2009). Determinación y complementación del nombre propio en español. *Nueva Revista de Filología Hispánica*, 52(2), 441–488.
- Epstein, R. (1993). The definite article: Early stages of development. In J. van Marle (Ed.), *Historical linguistics. Papers from the 10th International Conference on Historical Linguistics* (pp. 111–134). Amsterdam: John Benjamins.
- Epstein, R. (1994). *Discourse and definiteness: Synchronic and diachronic perspectives* (Unpublished doctoral dissertation). University of California San Diego, San Diego, CA.
- Epstein, R. (2001). The meaning of definite articles in cross-linguistics perspective. In E. Németh (Ed.), *Cognition in language use: Selected papers from the 7th International Pragmatics Conference* (Vol 1; pp. 174–189). Antwerp: International Pragmatics Association.
- Fernández Leborans, M. J. (1999). El nombre propio. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 1; pp. 77–128). Madrid: Espasa.
- García Gallarín, C. (1983). El artículo ante nombre propio de persona (Unpublished doctoral dissertation). Universidad Complutense de Madrid, Madrid, Spain.
- Gary-Prieur, M-N. (1994). *Grammaire du nom propre*. Paris: Presses Universitaires de France.
- Glare, P. G. W. (Ed.) (1982). *Oxford Latin dictionary*. New York, NY: Oxford University Press.
- Keniston, H. (1937). *The syntax of Castillian prose. The sixteenth century*. Chicago, IL: The University of Chicago Press.
- Koch, P., & Osterreicher, W. (2007). *Lengua hablada en la Romania: Español, francés, italiano*. Madrid: Gredos.
- Laca, B. (1999). Presencia y ausencia de determinante. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 1; pp. 891–928). Madrid: Espasa.
- Lapesa, R. (2000 [1961]). Del demostrativo al artículo. In *Estudios de morfosintaxis histórica del español* (pp. 360–387). Madrid: Gredos.
- Leonetti, M. (1999). El artículo. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 1; pp. 787–890). Madrid: Espasa.
- Longobardi, G. (1994). Reference and proper names: A theory of N-Movement in syntax and logical form. *Linguistic Inquiry*, 25(4), 609–665.
- Lyons, Ch. (1999). *Definiteness*. Cambridge: Cambridge University Press.
doi:10.1017/CBO9780511605789
- Ortiz Ciscomani, R. M. (2009). La creación y generalización del artículo definido. In C. Company Company (Ed.), *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal* (pp. 271–386). Ciudad de México: Fondo de Cultura Económica/UNAM.
- Ortiz Ciscomani, R. M., & Reynoso Noverón, J. (2012). La determinación y el nombre propio. Un estudio histórico de pragmática social en español. In E. Montero Cartelle & C. Manzano Rovira (Eds.), *Actas del VIII Congreso Internacional de Historia de la Lengua Española* (pp. 2313–2323). Santiago de Compostela: Meubook.

- Putzu, I., & Ramat, P. (2001). Articles and quantifiers in the Mediterranean languages: A typological-diachronic analysis. In W. Bisang (Ed.), *Aspects of typology and universals* (pp. 99–132). Berlin: Akademie.
- Real Academia Española. (2005). *Diccionario panhispánico de dudas*. Madrid: Santillana.
- Real Academia Española/Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.
- Reynoso Noverón, J. (2008). El género textual y la sintaxis del nombre propio. Estudio histórico. In C. Company Company & J. G. Moreno de Alba (Eds.), *Actas del VII Congreso Internacional de Historia de la Lengua Española* (pp. 2119–2129). Madrid: Arco/Libros.
- Rigau, G. (1999). La estructura del sintagma nominal: Los modificadores del nombre. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 1; pp. 311–362). Madrid: Espasa.
- Roca Urgell, F. (1996). *La determinación y la modificación nominal* (Unpublished doctoral dissertation). Universitat Autònoma de Barcelona, Barcelona, Spain.
- Selig, M. (1992). *Die Entwicklung der Nominaldeterminanten im Spätlatein. Romanischer Sprachwandel und lateinische Schriftlichkeit*. Tübingen: G. Narr.
- Trager, G. L. (1932). *The use of the Latin demonstratives (especially Ille and Ipse) up to 600 A.D., as the source of the Romance article*. New York: Publications of the Institute of French Studies.
- Traugott, E. C. (1995). Subjectification in grammaticalization. In D. Stein & S. Wright (Eds.), *Subjectivity and subjectivisation: Linguistic perspectives* (pp. 31–54). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511554469.003
- Traugott, E. C. (2003). From subjectification to intersubjectification. In R. Hickey (Ed.), *Motives for language change* (pp. 124–139). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486937.009
- Traugott, E. C. (2010). (Inter)subjectivity and (inter)subjectification: A reassessment. In K. Davidse, L. Vandelanotte & H. Cuyckens (Eds.), *Subjectification, intersubjectification and grammaticalization*. (pp. 29–74). Berlin: De Gruyter Mouton. doi:10.1515/9783110226102.1.29
- Traugott, E. C., & Dasher, R. B. (2002). *Regularity in semantic change*. Cambridge: Cambridge University Press.

Corpus

- Bize, M. (2010). *La vida de los peces*. Santiago de Chile: Ceneca Producciones. Film.
- DLNE = Company Company, C. (1994). *Documentos lingüísticos de la Nueva España (1525–1816). Altiplano Central*. Ciudad de México: Universidad Nacional Autónoma de México.
- Libro de buen amor* = Juan Ruiz, (1330–1343), *Libro de buen amor*. J. Corominas (Ed.). Madrid: Gredos, 1973.
- Maradona, D. (2000). *Yo soy el Diego*. Buenos Aires: Planeta.
- Probanza contra Alonso Zambrana, vecino de Baños de Graena (Granada), sobre injurias*. 1779. Archivo de la Real Chancillería de Granada. 10563/3. Manuscript.
- Probanza contra Bartolomé Santiesteban, vecino de Vélez Málaga, por heridas*. 1755. Archivo de la Real Chancillería de Granada. 10491/8. Manuscript.
- Probanza sobre injurias: Juan Pérez contra Miguel de Moya, vecinos de El Marchal (Granada)*. 1785. Archivo de la Real Chancillería de Granada. 10730/11. Manuscript.

- Provisión al corregidor de Loja, a pedimento de D^a. Claudia Pérez de Herrera*. 1782. Archivo de la Real Chancillería de Granada. 5266/196. Manuscript.
- Querrela criminal entre vecinos de Fiñana (Almería)*. 1746. Archivo Histórico de Almería. 43510/515. Manuscript.
- Real Academia Española. *Corpus de referencia del español actual (CREA)*. Madrid: RAE. <<http://www.rae.es>>
- Real Academia Española. *Corpus diacrónico del español (CORDE)*. Madrid: RAE. <<http://www.rae.es>>

Stylistic fronting in Old Spanish texts

Javier Elvira

Universidad Autónoma de Madrid

This paper provides new evidence on Stylistic Fronting in Old Spanish. The data show significant differences in the distribution and restrictions of Stylistic Fronting in main and subordinate clauses. They also provide new evidence about the syntagmatic split of phrases occurring simultaneously in many cases of Stylistic Fronting. I also address the question of whether this kind of fronting may be viewed as an instance of the emphatic movement or focus. Looking at textual evidence, the focal or expressive character of these expressions is not always easy to prove. These data lead to the conclusion that Stylistic Fronting should be viewed as a by-product or an epiphenomenon derived from multiple causes and motivations.

Keywords: Stylistic Fronting; Spanish; word order; main clause; subordinate clause; focus; movement

1. Introduction

In this paper, I discuss a peculiar word order pattern frequently found in Old Spanish texts, arguably the result of a movement called Stylistic Fronting (SF), which shifts a post-verbal constituent (usually an adjective, a participle or an infinitive) to the left of the finite verb. This movement also causes the presence of a subject gap such as that found in subject relatives and also in null-subject structures.

This structure was initially found in Icelandic and other Scandinavian languages, but also in other Germanic and Romance languages. The attribute 'stylistic' refers to the fact that the alleged movement seems to be optional, although this is somewhat debatable. This phenomenon differs from topicalisation in many respects, most importantly because it shows complementary distribution with a topical subject in canonical or preverbal position (cf. Eide, in this volume), and also because the type of fronted lexical material is different from the one that can be found in topicalisation movements. It may also be attested in subordinate clauses.

The first scholar who paid attention to this movement was Joan Maling in her work “Inversion in embedded clauses in Modern Icelandic” (1980, 1990), in which she studied a special type of movement through which predicative elements of Icelandic are moved into a preverbal position. The research that followed Maling’s papers has shown that SF was highly productive at earlier stages of other Scandinavian languages such as Old and Middle Danish (Hrafnbjargarson, 2004a, 2004b), Old Swedish (Platzack, 1988), but less frequent in Modern Insular Scandinavian (e.g., Modern Icelandic, Faroese; Maling, 1980, 1990).

We also know that similar but not necessarily identical facts were frequent in some Romance varieties, such as Old Italian (Franco, 2009), Old French (Mathieu, 2006) and Old Catalan (Fischer, 2005) and that they are less productive (or rather unproductive) in modern Romance languages. We can find some references to SF in Old Spanish in Fontana (1993, pp. 65–66) and more recently in Fischer (2010, 2014), but the evidence provided by these authors is not systematic or exhaustive.¹

Most of the efforts in the study of SF have been focused on its discourse effects, specifically on its relation to the different types of focus and other discourse motivated movements. Another bulk of research has been developed in a generative or minimalist framework and has tried to investigate SF at the interface of syntax and pragmatics, aiming to find out what kind of movement could give rise to SF, what kind of syntactic element undergoes this movement and what is the target or landing site of this movement.

SF shows a significant variation among different languages in which it has been identified, both in the syntactic distribution of the facts and with relation to the elements or categories affected by this movement. The heterogeneous character of the data hinders a uniform definition and understanding of the phenomenon.

The present work tries to provide new empirical evidence about SF in Old Spanish and at the same time attempts to understand this evidence in light of the evolutionary trends of word order from Latin to Romance languages.

2. Textual evidence

After providing a definition of SF that is broad enough to match the data from languages as different as those in which the process has been documented, we must

1. Some of the facts that Fernández-Ordóñez (2008–2009, pp. 14–19) brings together under the label *inversión predicativa* ‘predicative inversion’ may be viewed as examples of SF.

now look at the texts in search of empirical information about its function in Old Spanish. The data for this study come from the exhaustive screening of some texts of Alphonse X and the search for occurrences of specific word order patterns in the texts included in the CORDE digital database.

As previously mentioned, SF has been described as a syntactic movement that places a predicative element in a preverbal position, adjacent to the verb (however, see Section 5 for exceptions to this), that is unusual or not frequently attested for this element. The items that are frequently found in this construction in Old Spanish are predicative adjectives and adverbs used with copulative verbs and also participles and infinitives used in verbal periphrases:

- (1) *Contado auemos [...] / Mucho es [...] / Fechas son [...]*
 told have.1PL / much is.3SG / done are.3PL
Demandar quiere [...]
 demand.INF want.3SG

The items that undergo this process are not usually available to other movements, specifically topicalisation, which moves entire constituents to a preverbal position. On the contrary SF is a movement that takes place within a verbal phrase and only affects some of the elements that are integrated in it.

A data search in Old Spanish texts provides some evidence that many constructions exist that seem to fit the above description. In fact, we can find adjectives or adverbs functioning as predicatives of copulative verbs which are located to the left of the verb:

- (2) a. *Ca en las cosas en las quales non ay manifesta*
 since in the things in which not there-is.3SG manifest
perdición nin ajenamiento de Dios, bueno es obedesçer
 downfall nor remoteness from God good is.3SG to-obey
al padre
 to-the father
 ‘Since in the things where there is no obvious perdition or remoteness from God, it is a good thing to obey the Father’
 (*Barlaam e Josafat*, ms. S, p. 437, CORDE)
- b. *E por ende sy bien estan los Reys non se*
 and moreover if well are.3PL the Kings not themselves
deuen mouer dende
 must.3PL to-move away
 ‘Moreover, if the kings are well, they should not move away’
 (*Castigos*. BNM ms. 6559, 186r, CORDE)

Infinitives governed by main verbs (a), participles in verbal periphrases (b) and predicative elements (c) may also be located in this preverbal position:

- (3) a. *Obedecerle ssopo muy bien; ca nunca ffizo cosa que*
to-obey-to-him knew.3SG very well as never did.3SG thing that
contra la ffe de Ssanta Iglesia de Rroma ffuese
against the faith of Holy Church of Rome was.3SG
'He knew how to obey him very well, as he never did anything against
the faith of the Holy Church of Rome' (Setenario, p. 14, CORDE)
- b. *Bendicho es el que uiene en el nombre de*
blessed is.3SG he who come.3SG in the name of
nuestro señor
our Lord
'Blessed is the one who comes in the name of our Lord'
(Estoria de España I, 87r)
- c. *Cancro es llamado el quarto ssigno, que quier tanto*
Cancro is.3SG called the fourth sign which wants also
decir en nuestro lenguaie commo cangreio
to-say in our language as crab
'The fourth (zodiac) sign is also called Cancer, which means "crab" in
our language' (Setenario, p. 96, CORDE)

SF in Old Spanish always takes place in positions adjacent to the verb. The only element that may be located between the verb and the fronted item is the clitic pronoun:

- (4) *e esto de como fue todo, contado lo auemos en el*
and this on how was.3SG everything told it have.1PL in the
primero libro de los Macabeos
first book of the Maccabees
'and how everything was, we have told it in the first book of the Maccabees'
(General Estoria IV, 4)

There are other syntactic features that have been frequently mentioned in the descriptions of SF. Some of these may also be found in Old Spanish, as shall be shown in the following.

2.1 Subject gap

Some researchers of other languages have observed that SF is possible if there is a subject gap in the clause, i.e. if there is no overt subject in the initial subject position (Maling, 1980, p. 181). Old Spanish data seem to confirm this idea, as SF never occurs with an explicit initial subject. Sometimes there is no subject at all:

- (5) a. *Dicho es en el pueblo: Si dexare ell omne su mugier [...] wife*
 said is.3SG in the people if left.3SG the man his wife
 ‘It is said by the people that if a man left his wife [...]’
 (General Estoria IV, 77r)
- b. *Dicho auemos ante desto en la estoria de los Reys de babilona de cómo [...] Babylon of how*
 said have.1PL before this in the history of the Kings of Babylon of how
 ‘We have said before in the history of the Kings of Babylon how [...]’
 (General Estoria IV, 54r)

However, cases of SF with overt subject are perfectly possible, specifically with the copulative verb *ser* ‘to be’, which usually appears with explicit subject:

- (6) *Mucho son honrrados los tus amigos, e mucho es confirmado el señorío dellos*
 much are.3PL honoured the your friends and much is.3SG confirmed the lordship of-them
 ‘Your friends are very honoured, and their lordship is confirmed’
 (Setenario, 28v, CORDE)

SF is especially frequent when the subject position is filled by a subordinate sentence in constructions that might be viewed as impersonal. A speaker’s preference for not placing ‘heavy’ subjects sentence-initially can help explain why SF obtains in these cases:

- (7) *sabido es que la compañía es nehesaria al onbre todos sus días [...]*
 known is.3SG that the company is necessary to-the man all his days
 ‘It is known that the company is necessary for man all his days’
 (Pedro de Toledo, *Guia de los perplejos de Maimónides*. BNM ms. 10289, CORDE)

A well-known variant of this construction is the archaic expression *conocida cosa sea que* (‘it must be known that’), and similar uses:

- (8) *conosçida cosa es que, segund la flaqueza humana, más inclinados somos a errar que a bien obrar*
 known thing is.3SG that according the frailty human more prone are.1PL to err than to well work
 ‘It is known that, according to human frailty, we are more prone to error than to doing right’
 (*Crónica de Don Álvaro de Luna*, 248, CORDE)

2.2 Initial position in main sentences

According to my data, it seems clear that SF is especially linked to the initial position in main sentences. Indeed, many of the Spanish SF examples occur at the absolute beginning of the sentence (9a), sometimes after a pause or break (we mark this break with the sign #) (9b). They are also possible after *et* ‘and’, *ca* ‘since’ and *pues* ‘since’, and other connective particles (9c):

- (9) a. *partidos son por la sapiencia de dios et mudara*
 left are.3PL by the wisdom of God and will-change.3SG
ell los tiempos et los dias de las fiestas dellos
 he the times and the days of the holidays of-them
 ‘[The periods of time] have been divided by the wisdom of God and
 he will change the dates and the days of their holidays’
 (General Estoria IV, 33)
- b. *mas si lo recebiere, # tenido es de finir*
 but if it will-receive.3SG had is.3SG of to-finish
el pleito
 the lawsuit
 ‘but, if he received it, he is expected to finish the lawsuit’
 (Vidal Mayor, 138 CORDE)
- c. *ca dicho es assaz como Sem fizo a Arfaxat*
 so said is.3SG enough how Sem did.3SG to Arfaxat
 ‘so it is frequently told what Sem did to Arfaxat’
 (General Estoria I, 66r)

2.3 Subordinate sentences

In Old Spanish, SF is possible in main clauses, as in many of the above-quoted examples, but it may also occur in subordinate clauses. Relative clauses in Old Spanish are a possible context for SF (10a) but it is also possible in other subordinate clauses such as complement (10b) or adverbial clauses. The sentence quoted in (10c) combines two instances of SF occurring in complement and conditional clauses:

- (10) a. *passadas las cosas que dichas auemos en el regnado*
 passed the things that said have.1PL in the kingdom
del Rey Euilmoradac murio esse Rey
 of-the King Euilmoradac died.3SG that King
 ‘After the things we have told occurred in the kingdom of Euilmoradac,
 this king died’
 (General Estoria IV, 49r)

- b. *dize en el postremero capitulo de las razones dela quinta*
 say.3SG in the last chapter of the issues of-the fifth
edad que tres fueron las catiuidades de los fijos
 age that three were.3PL the kidnappings of the children
de israhel
 of Israel
 ‘It is said in the last chapter of the issues of the fifth age that there were
 three kidnappings of the children of Israel’ (General Estoria IV, 1r)
- c. *Dixol essora el Rey: Se lo yo que destroyda*
 said.3SG-him. then the King know.1SG it I that destroyed
a de ser si uerdad es lo que los adeuinos dizen
 have.3SG of to-be if truth is.3SG what the seers say.3PL
 ‘Then the king said: I know that it must be destroyed, if it is true what
 the seers say’ (General Estoria IV, 15r)

3. Focus meaning?

Some researchers of SF consider that focus or another discourse emphatic factor is involved in the use of SF (Fischer & Alexiadou, 2001; Franco, 2009; Hrafnbjargarson, 2004a, 2004b). In contrast, other linguists support the idea that SF is triggered by some syntactically motivated movement alone (Maling, 1980, 1990; Molnar, 2010, p. 42). In the following we will try to answer whether there is a focal meaning in these Old Spanish constructions. Can we consider that this kind of fronting is simply an instance of the emphatic movement that we usually consider under the name of focus? Some researchers have suggested that some cases of predicative fronting can be viewed as instances of focus (Fernández-Ordóñez, 2009, n° 2.3). Indeed, the context may sometimes help us perceive the semantic emphasis, which is typical of focus:

- (11) *E sy bueno era a cavallo, fuerte e fiero fue quando*
 and if good was.3SG on horse strong and fierce was.3SG when
se vio a pie
 himself saw.3SG on foot
 ‘And if he was good on a horse, he was also strong and fierce when he was
 on foot’ (Cuento muy fermoso de Otas de Roma, 42, CORDE)

In (11) we observe a clear contrast between *a cavallo* ‘on a horse’ and *a pie* ‘on foot’, which reinforces the semantic contrast between *bueno* ‘good’ and *fuerte e fiero* ‘strong and fierce’, which are also foci. However, most of the instances of SF we

find in Old Spanish texts cannot be considered such clear cases of contrastive or informational focus:

- (12) a. *e el nuestro tuerto, sabido es por el mundo*
 and the our mistake known is.3SG around the world
 ‘and our error is well-known around the world’
 (General Estoria II, 180v)
- b. *E por lo que vos dicho avedes seguramente podemos dezir*
 and from what you said have.2PL certainly can.1PL to-say
que movimiento de grand locura nos ha traído aquí
 that movement of great madness us has.3SG brought here
 ‘And from what you have said, we can deduce that an attack of great
 madness has brought us here’
 (García de Salazar, *Istoria de las bienandanzas e fortunas*, CORDE)

In cases like the ones above, it is difficult to perceive a clear contrast effect in the fronted element (e.g., *sabido y no olvidado* ‘known and not forgotten’, *avedes dicho y no escrito* ‘you have said rather than written’, etc.).

At the same time, no Modern Spanish construction can be strictly viewed as similar to the uses described above. The most similar example that we can find in present-day Spanish is what Leonetti and Escandell Vidal (2009) call *verum focus*, which is present in sentences like (13):

- (13) *Algo debe saber Juan / Bien lo sabe Dios*
 something must.3SG to-know Juan / well it know.3SG God
 ‘John must know something’/‘God knows it very well’

These sentences carry some kind of polarity focus, i.e. they emphasise the whole sentence rather than one of its constituents:

- (14) a. *Algo debe saber Juan = Juan sabe algo,*
 something must.3SG to-know Juan = Juan know.3SG something
por supuesto que sí
 of course that yes
 ‘John knows something (of course)’
- b. *Bien lo sabe Dios = Dios lo sabe muy bien, sin*
 well it know.3SG God = God it know.3SG very well without
duda alguna
 doubt some
 ‘God knows it very well, without a doubt’

The idea that our examples of SF may carry *verum focus* is not always easy to prove. We lack evidence from prosody and intonation that would allow us to

support this assumption clearly. Indeed, the emphatic reading may sometimes be the effect rather than the cause of SF. Obviously these expressions are to some extent marked formally, and that status may frequently lead to a parallel expressive reading of some of these examples. The emphatic reading, however, may be more or less clear depending on the specific cases and their context. On the other hand, when dealing with this issue, it is not always easy to rid ourselves of our grammatical feelings as contemporary speakers of Spanish, which may lead us to find emphasis in a construction that probably had no such reading.

In view of the difficulties in ascertaining the focal character of these constructions in all circumstances and with enough certainty, we have decided not to consider this side of the issue further, while at the same time not excluding the possibility that some of these constructions may have a focal or emphatic reading. Rather, we choose to look into our textual data in an attempt to find any regular grammatical factor that may promote SF, under the assumption that even if there is a pragmatic factor in many cases, we are allowed to assume that SF is always grammatically motivated.

4. Stylistic fronting in main clauses

We have already shown that SF in main clauses is usually possible when the verb phrase is located in absolute initial position or when it is followed by some connectives, like *et*, *ca* or *mas* (15a). It is also possible when this initial position is preceded by a topical element or an expression performing the role of introducing contextual or background information (15b):

- (15) a. *Contado auemos nos ante desto como fueron seys los*
 told have.1PL we before this how were.3PL six the
Reys que en babilonna e en caldea regnaron
 kings that in Babylon and in Chaldea reigned.3PL
 ‘We have already told before that there were six kings who reigned in
 Babylon and in Chaldea’ (Estoria de España IV, 47r)
- b. *Quien de su enemigo ha piedat, #digno es de*
 who of his enemy has.3SG mercy worthy is.3SG of
grand bondat
 great kindness
 ‘He who has mercy on his enemy is worthy of great kindness’
 (C. Sánchez de Vercial, *Libro de los exemplos por A.B.C.*, 24, CORDE)

This very first position after an initial topic followed by an intonation break, is usually reserved for the verb in many Old Spanish texts (Elvira, 1993):

- (16) *Et desdeque anolescio que cerraua ya mi puerta en*
 and since got-dark.3SG that closed.1SG then my door in
tiniebra, # llegaron ellos et salieron de casa entonces
 darkness arrived.3PL they and left.3PL of home then
 'And when it got dark and I was closing my door in darkness, they arrived
 and left home then' (General Estoria II, 6v)

Due to the fact that the verb usually follows the topic or background information, languages that adopt this word order pattern are sometimes called TVX languages, that is Topic + Verb + other constituents (if any). Vennemann (1974) argues that this arrangement corresponds to an intermediate state between Latin word order and the new SVO order of the Romance languages. Those who assume the existence of this intermediate type usually suppose that the new SVO order appeared firstly as a specific implementation of the TVX model, that is, they are constructions with a topical subject. The frequency of occurrence of constituents other than the subject in the topical initial position becomes increasingly lower and therefore the SVO type has become progressively more frequent until it eventually turned into the unmarked subject position.

Nevertheless, recent typology classifies those languages that tend to locate the verb in second position, after any other sentence constituent, as V2. Some Germanic languages of past or present times belong to the V2 type. In fact, many researchers of SF in other languages have considered the hypothesis that SF takes place in V2 languages, when the initial preverbal position is empty. The need to fill this gap forces the shift of the element located after the verb to the initial position. This helps explain the subject-gap condition that many scholars have detected in SF in other languages.

This explanation is very interesting, as it suggests that SF is caused by the absence of any element to the left of the verb. In any case, from an empirical point of view we can say that the Old Spanish verb is not systematically placed in second position and may also appear in other positions. In fact, the evidence of V1 in Old Spanish texts is quite abundant, especially in the narrative prose of Alfonso X. The initial position for the verb is frequent in those cases in which there are no elements with marked topicality. This initial position of the verb in the absence of a topicalised fronted element may also be viewed as consistent with the (T)VX word order pattern:

- (17) a. *Salieron las madres et las mugieres et las mancebas*
 left.3PL the mothers and the women and the girls
empos ellos
 after them
 ‘The mothers and the women and the girls left after them’
 (*General Estoria* IV, 53v)
- b. *Murio Josaphat et fue enterrado con sus padres en*
 died.3SG Josaphat and was.3SG buried with his parents in
la Çiupdat de dauid
 the city of David
 ‘Josaphat died and he was buried with his parents in the city of David’
 (*Biblia. Escorial* I.j.8, 52, CORDE)

However, the Old Spanish preference for TVX affects SF in a specific way. As mentioned above, among the items directly affected by this fronting are predicative elements that are linked to copulative or auxiliary verbs. These are verbal elements with a significant grammatical role and a relatively reduced lexical meaning. However, these verbs show a specific behaviour in relation to word order because they usually avoid the first position in main clauses. Indeed, initial or fronted position (XV) for copulative or auxiliary verbs like *ser* ‘to be’ and *aver* ‘to have’ is unusual in many Old Spanish texts. This fact was observed a long time ago by the pioneers of the research into Romance linguistics (Meyer-Lübke, 1900, § 724) and Old Spanish (Menéndez Pidal, 1976a, 1976b) and has been confirmed by recent research (Rodríguez Molina, 2010, p. 1362).

Additionally, Alfonso X’s *General Estoria* provides evidence that the verb *ser* is uncommon in initial sentence position or right after the initial topical periphery, especially in the third present person form *es*. On the contrary, the form *es* is usually preceded by some other element, among others, the subject of the clause (18a), an adverb, including *non* and *nin* (18b)–(18c), a subordinating conjunction (18d) or a connective element like *et*, *ca* or *mas* (18e):

- (18) a. *Et ell amor es cosa creediza [...]*
 and the love is.3SG thing credible
 ‘and love is something credible [...]’
 (*General Estoria* II, 65r)
- b. *e assi es aquí e en todas otras cosas tales*
 and like-this is.3SG here and in all other things such
como esta
 as this
 ‘and like this is here and in all other things’
 (*General Estoria* II, 291v)

- c. *Et non es aquel el fijo de Gedeon que fuxo*
and not is.3SG that the son of Gideon who fled.3SG
'And it is not the son of Gideon, who fled' (General Estoria II, 294v)
- d. *Et maguer que es cosa espantosa et enoiosa, alli*
and although that is.3SG thing frightening and annoying there
morare yo contigo
will-live.1SG I with-you
'And although it is a frightening and annoying thing, I will live there
with you' (General Estoria II, 339V)
- e. *Ca es ya muy grand la nuestra desonrra [...]*
since is.3SG now very big the our disgrace
'Since our disgrace is now very great [...]' (General Estoria II, 42v)

The same is true about the verb *aver* 'to have', which systematically avoids sentence-initial position. The fact that the number of syllables may influence the position of the verb *ser* could be related to the impossibility of unstressed forms beginning a sentence. Indeed, the persistence of the Tobler-Mussafia law with respect to the behaviour of unstressed forms in Old Romance is well known. Given this, we might think that monosyllabic forms of *ser* are phonetically weak and are thus avoided in initial position. However, it is noteworthy that some polysyllabic (and probably tonic) verbal forms are also excluded from initial position.

Particularly relevant in this regard is the three-syllable form *avemos* 'we have', which may also be included in a perfect tense periphrasis. In this compound tense, the auxiliary verb is usually located before the participle (e.g., *avemos dicho*, *avemos fecho*). However, just like unstressed forms, it avoids initial position and causes the fronting of the participle (19a). The same occurs after an initial topic (19b):

- (19) a. *Departido auemos de cómo Regno este*
told have.1PL of how reigned.3SG this
'We have told how he reigned [...]' (Estoria de España II: 31v)
- b. *Las razones [...] que lucano dixo, # contadas las*
the arguments that Lucano said.3SG told them
auemos aquí
have.1PL here
'The arguments [...] that Lucan made, we have given them here'
(Estoria de España I: 49r)

In summary, there are certain syntactic constraints that automatically determine the position of the verb (whether copulative, auxiliary or modal) and the predicative element that is linked to it. In the texts examined, the initial position, absolute

or after an initial topic, determines the immediate reversal of the usual relative ordering of both elements. In this context, SF is an immediate consequence of this circumstance, regardless of the conditions of focus or emphasis that may be present from the context.

It is worth remembering that the verbal elements that more frequently head constructions in which SF is undergone are copulative verbs, auxiliary and other verbs involved in verbal periphrases. In short, the elements involved in these constructions are usually verbs with little or no lexical but extensive grammatical content. We have already seen that purely grammatical elements (like present tense *es* and others) are excluded from initial position. The fact that these initial elements have lost some of their lexical meaning may be the reason that prevents them from being located in the first position, as required by the TVX pattern in Old Spanish, which always places a lexical verb in initial or post-topical position.

5. Split phrases and analytical futures

According to the data above, placing a complex verbal phrase in initial position has the secondary effect that the participle or predicative element of this verbal phrase is fronted. In addition to this, the complex phrase may sometimes be split by an unstressed pronoun, presumably following the Tobler-Mussafia law, which describes the preference of unstressed elements in Romance as leaning enclitically on the first element of the sentence (Castillo Lluch, 1996):

- (20) *Contado uos auemos ante desto en las razones del tiempo de*
 told you have.1PL before that in the reasons of-the time of
Othoniel juyz de israhel
 Othoniel judge of Israel
 ‘We have already told you about the events of the time of Othoniel, judge
 of Israel’ (General Estoria IV: 184v)

The splitting of phrases is a frequent phenomenon in Latin, probably due to the lower syntagmatic cohesion in that language. In Latin it affects both verbal and nominal phrases and is related in many cases to focus effects.² Adams (1994a, pp. 15–33; 1994b, p. 106) has shown that certain complex predicates can be split by the copulative or auxiliary verb:

2. These discontinuous structures were also possible in Classical Greek and have been traditionally labeled ‘hyperbaton’ (see Mathieu & Sitaridou, 2002).

- (21) *Catilina* [...] *certus erit* *competitor*
 (< *Catilina* [...] *erit certus competitor*)
 Catiline [...] real will-be.3SG competitor
 ‘Catiline will be a real competitor’ (Cic. Att. 111; Adams, 1994a, p. 16)

The splitting of phrases also occurs in Old Spanish. Slawormirski (1986, pp. 94–95) deals with this issue of hyperbaton (to which he gives the more traditional name of *tmesis*). This syntagmatic split can occur in Old Spanish complex nominal predicates, sometimes due to focus effect, as in Latin. Quantifiers, which usually have a rather emphatic meaning, are very prone to fronting (Camus, 2008) in combination with phrase splitting:

- (22) *et dixo que mucho era triste e se tenie*
 and said.3SG that very was.3SG sad and himself considered.3SG
por malandant
 for unlucky
 ‘and he said that he was very sad and considered himself unlucky’
 (General Estoria II, 289v)

However, in most cases, beyond the influence of focus, it is the fronting of a quantifier to initial position that favours this type of splitting:

- (23) a. *Mucho fueron marauillados de la grant mortandat que*
 very were.3PL amazed of the great mortality that
fallaron de los moros
 found.3PL of the Moors
 ‘They were very amazed by the great mortality they found in the Moors’
 (Estoria de España II: 253r)
- b. *Mucho era bien andant enneas en affrica con la*
 very was.3SG well going Aeneas in Africa with the
reyna dido
 Queen Dido
 ‘Aeneas was very fortunate with Queen Dido in Africa’
 (Estoria de España I: 26v)

Hyperbaton in Old Spanish is also possible with compound tenses and infinitive periphrases and is usually linked to initial or post-topical position:

- (24) a. *de como este Rey don Ramjro Regno, # dicho lo*
 on how this King don Ramiro reigned.3SG said it
auemos en el capitulo de ante deste
 have.1PL in the chapter of before of-this
 ‘We have said how King Ramiro reigned in the previous chapter’
 (Estoria de España II: 50v)

- b. *Obedecer -le ssopo muy bien; ca nunca ffizo*
 to-obey -to-him knew.3SG very well since never made.3SG
cosa que contra la ffe de Ssanta Iglesia de
 thing which against the faith of Saint Church of
Roma ffuese
 Rome was.3SG
 'He knew how to obey him very well, since he never did anything that
 was against the Church of Rome' (Setenario, p. 14, CORDE)

These data are reminiscent of the analytic forms of the future and conditional (referred to collectively herein as AFC) in Old Spanish, a well-known phenomenon in which the verbal stem is split from the ending of the future and conditional. These endings are former modal auxiliaries operating in periphrasis with the infinitive. The texts under scrutiny in this research show that AFC, like SF, are possible in main sentences and are also restricted to initial (25a) or post-topical position – TVX contexts – (25b) and (25c)–(25d):

- (25) a. *Dixol el: trabaia m -e dend*
 said.3SG-to-him he to-work me will.1SG on-this
quanto sopier
 as much know.1SG
 'He said to him: I will work with this the best I know how'
 (General Estoria IV: 4r)
- b. *dixo so padre a Badiza: poder m -ies tu guarescer*
 said.3SG his father to Badiza can me would.2SG you to-take-care
esta ninna
 this girl
 'His father said to Badiza: you could take care of this girl (for me)'
 (General Estoria IV: 4r)
- c. *si-m la guarescieres, # onrar t -e yo*
 if-me. it take.care.2SG to-honour you will.1SG I
 'If you were to take care of her (for me), I would honour you'
 (General Estoria IV: 4r)
- d. *Agora por que ell fue Rey tan poderoso [...], # contar*
 now because he was.3SG king so powerful [...] to-tell
uos -emos luego la su nacencia
 you will.1PL then the his birth
 'Since he was such a powerful king, we will now tell you of his birth'
 (General Estoria IV, 3r)

To a lesser extent, the parallelism between SF and AFC is also manifest in subordinate clauses, where the TVX order is unusual but possible. In these isolated cases, AFC may occur immediately after an internal topic:

- (26) *sabido es que las menores # levar -las ha*
 known is.3SG that the minors to-take them will.3SG
cuesta ayuso
 slope upwards
 ‘It is known that he will take the minors uphill’

(Pérez de Guzmán, *Poesías*, 434, CORDE)

Fontana (1993, p. 84) noticed this similarity between AFC and SF and wrote that ‘split futures also manifest the fronting’. Given the data above, this similarity also extends to the syntactic distribution of split verbal phrases in general, so that, ultimately, the AFC can be considered a specific case of a more general split phrase that was frequent in Old Spanish. In turn, Bouzouita (2011, 2012) has noted that the syntactic distribution of AFC fits perfectly with the general trends of clitic placement.³ We therefore find a group of relatively heterogeneous syntactic phenomena (split phrases, AFC, enclisis and SF) which share a similar syntactic distribution.

6. Stylistic fronting in subordinate clauses

The above data show a preference for initial position by split verbal phrases. This notwithstanding, SF in subordinate clauses is quite unusual in Old Spanish, although it is perfectly possible:

- (27) a. *Mager que dicho sea que non vala otor si*
 although that said was.3SG that not is-worth.3SG warranty if
non fuere dado en Vbeda [...]
 not was.3SG given in Úbeda
 ‘Although it was said that the warranty is not valid if not given in Úbeda [...]’
 (1251, *Fuero de Úbeda*, p. 390, CORDE)
- b. *fuy presente a esto que dicho es e con los*
 was.3SG present at this that said is.3SG and with the
dichos testigos
 said witnesses
 ‘I was present at what was said along with the above-mentioned witnesses’
 (1384, *Acta de reconocimiento*, p. 129, CORDE)

3. Unstressed pronouns in Old Spanish functioned as enclitics leaning to the verb in initial or post-topical position (Castillo Lluch, 1996, 2002).

However, the list of elements that may be placed before the verb in subordinate clauses is somewhat longer than the list of elements that may undergo SF in main clauses. Indeed in subordinate clauses not only may predicative elements (adjectives, adverbs, participles, etc.) be found before the verb, thus undergoing SF, but other major constituents or entire phrases can also be moved to a preverbal position, such as direct objects (28a), adverbial prepositional phrases (28b)–(28c), passives agents (28d), etc.:

- (28) a. *ninguna cosa que pena corporal enduziesse non la*
 no thing that punishment corporal cause.3SG not it
queríamos ditar nin poner en este libro
 wanted.1PL to-dictate nor to-put in this book
 ‘We do not want to dictate nor put into this book anything that may cause bodily harm’ (1250, *Vidal Mayor*, p. 12, CORDE)
- b. *Et ssi ssacar armas o conducho o otra cosa que en*
 and if took-out.3SG arms or food or other thing that in
casa del rey ssea que lo peche assy como
 house of-the king is.3SG that it pay.3SG such as
de suso dixiemos
 above said.1PL
 ‘And if he took arms, food or another thing that was at the king’s house, he must pay for it, as we have said above’
 (*Espéculo*, p. 14, CORDE)
- c. *et periglo sería tenido de fazer los instrumentos de*
 and danger would-be.3SG had of to-make the instruments of
las notas que con la su propia mano tomó
 the note-taking that with the his own hand took.3SG
et escriuió
 and wrote.3SG
 ‘and it would be considered dangerous to have to make the instruments of note-taking that he would use and would write with himself’ (1250, *Vidal Mayor*, p. 237, CORDE)
- d. *Enpero no es segura cosa que por clérigos ordenados*
 however not is.3SG sure thing that by priests ordained
sean feitos los públicos instrumentos
 are.3PL made the public instruments
 ‘However, it is not sure that the public instruments are made by ordained priests’ (1250, *Vidal Mayor*, p. 239, CORDE)

On the other hand, the observed blocking of SF by negative adverbs does not hold in subordinate clauses. This observation is probably not trivial, as it suggests that

the preverbal position is not essential to understand the movement of elements of the verbal phrase in subordinate clauses:

- (29) a. *poco uso dezimos aqui sobresto, non por que sabudo*
 little use say.1PL here on-that not because known
non sea
 not is.3SG
 ‘We devote little space to that, not because it is not to be known [...]’
 (General Estoria II, 22r)
- b. *nin yremos a otra tierra daqui adelant; que mas*
 nor will-go.1PL to another land from-here on which more
e mejor temprada non sea que esta
 and better moderate not is.3SG than this
 ‘From now on we won’t go to another country that is more moderate
 and has a better climate than this one’ (General Estoria II, 83r)
- c. *De guisa que el que y fuere; que nunca pobre sea*
 of manner that he who there go.3SG that never poor is.3SG
 ‘So that whoever were to go there, would never be poor’
 (General Estoria II, 265v)

Instances of multiple fronting of prepositional phrases are not difficult to find. In these cases, the verb is located in absolute final position:

- (30) *Et si nos lo fezimos por quebrantar la ley [...] nin por fazer*
 and if we it did.1PL for to-break the law nor for to-do
con el ninguna cosa que contra el nuestro sennor nin
 with him no thing that against the our Lord nor
contra la su ley nin la su creencia nin contra la
 against the his law nor the his belief nor against the
su onrra sea
 his honouris.3SG
 ‘And if we did not do it to break the law nor do with him anything that is
 against our Lord, his law, his belief or his honour’ (General Estoria II, 79r)

Several years ago, in an earlier paper on pronominal clitic position (Elvira, 1987), I observed that the subordinate verb in some Old Spanish texts showed a residual tendency towards final position. This tendency was never systematic or statistically dominant but was easy to find. These data helped me explain the fact that interpolation of clitic pronouns is possible only in subordinate clauses:

- (31) *e fazen se non uesibles deguisa que los non*
 and make.3PL- themselves not visible so that them not
puede ueer ninguno
 can.3SG to-see no-one
 ‘And they turn invisible so that no one can see them’

(*General Estoria* IV, 3r)

Final-verb structures in subordinate clauses are also possible in Old Spanish, though these patterns are unusual and residual in main clauses:

- (32) *en manera que nunca pueda auer el offitio que*
 in way that never can.3SG to-have the job that
por su culpa perdió ni otra cosa que ad aqueill
 because-of his fault lost.3SG nor other thing that to that
offitio semeillasse
 job was-similar.3SG

‘So that he may never have the job he lost because of his fault or another thing similar’

(1250, *Vidal Mayor*, p. 141, CORDE)

- (33) *nin podría sser ffirmе ninguna cosa que ssobre tal*
 nor could.3SG to-be firm no thing that about such
rrazon fuesse
 reason was.3SG

‘Nor could anything that was about this reason be firm’

(*Setenario*, p. 48, CORDE)

I am not the first to write about archaic OV patterns in an Old Romance language. Other classical researchers, like Gerhard Rohlfs (1949–1954, III, p. 217f.), noticed that in Old Italian and Old French the auxiliary may be located after the participle and that this order was especially frequent in relatives and other subordinate clauses. In turn, Bauer (2009, p. 249) said that the tendency toward OV order is stronger in subordinate clauses than in main clauses and mentions the existence of verb-final patterns in Umbrian, Oscan, and Old French.

According to these data, it is conceivable that SF, which is scarce in subordinate clauses in Old Spanish, can be understood as a particular effect of a larger trend of the verb in which it is positioned at the end of the sentence. This trend can be considered a historical residue and may give rise to different syntactic phenomena that are unknown in main clauses, such as the interpolation

of clitic pronouns or the position before the verb of very different syntactic elements.

7. Summary and conclusions

SF is a rather heterogeneous phenomenon, which is difficult to describe and define in a single way that may be appropriate for every language. It is diffuse, when viewed from a Romance perspective, and even more elusive when considered from a broader perspective. In any case, SF seems to be present in Old Spanish. We presented some data that may be viewed as similar to other cases which have been described as SF in other languages. SF may appear in main and also in subordinate clauses, but there seem to be different motivations for SF in either context.

In main clauses, SF is essentially linked to initial or post-topical position and, thus, is conditioned syntactically. This syntactic influence on the phenomenon is even greater due to the verbs having developed grammatical functions (copulative, auxiliary, modal, etc.) usually being excluded from initial position. This observation allows us to reverse the definition of SF, since it would be more accurate to talk about postposition of the verb rather than fronting of the predicative element, while still considering the same phenomenon.

SF may be accompanied by the splitting of the phrase by an unstressed element, usually a pronoun. This allows us to highlight the parallel between SF and AFC, which are also, ultimately, processes of syntagmatic splitting. Both phenomena are the result of the interference of enclisis of the unstressed pronouns, which is also related to the initial or post-topical position.

In subordinate clauses, SF is more scarce but possible and can be said to be related to a sporadic preference of the verb to be preceded by other syntagmatic elements of a different nature and size. These cases are occasional and difficult to systematise and can be put in relation to some Romance residues of the old verb-final order, which was maintained over a longer period in subordinate clauses.

The possibility that there may be a focus motivation for these data is difficult to determine, as it is not always easy to perceive the expressive factors clearly or the type of focus that motivates the appearance of the phenomenon. The evidence that SF is a direct consequence of some other purely syntactic factors is hardly compatible with the discourse-related understanding of the examples. Ultimately, many cases of focal reading might be the effect rather than the cause of the phenomenon.

SF is, to some extent, an elusive phenomenon, that can be quite difficult to grasp, and may be analysed as the secondary effect or an epiphenomenon derived from multiple causes and motivations (TVX, clitic laws, final verb residues). As these causes eventually disappeared, SF was also lost forever from Spanish grammar.

References

- Adams, J. N. (1994a). *Wackernagel's law and the placement of the copula esse in Classical Latin*. Cambridge: The Cambridge Philological Society.
- Adams, J. N. (1994b). Wackernagel's law and the placement of unstressed personal pronouns in Classical Latin. *Transactions of the Philological Society*, 92(2), 103–178. doi:10.1111/j.1467-968X.1994.tb00430.x
- Bauer, B. L. M. (2009). Word order. In Ph. Baldi & P. Cuzzolini (Eds.), *New perspectives on historical Latin syntax*, Volume 1: *Syntax of the Sentence* (pp. 241–316). Berlin: Mouton de Gruyter.
- Bouzouita, M. (2011). Future constructions in Medieval Spanish: Mesoclisys uncovered. In R. Kempson, E. Gregoromichelaki & C. Howes (Eds.), *The dynamics of lexical interfaces* (pp. 91–132). Stanford CA: CSLI.
- Bouzouita, M. (2012). Los futuros en la *Fazienda de Ultra Mar*, In E. Montero Cartelle & C. Manzano Rovira (Eds.), *Actas del VIII Congreso Internacional de Historia de la Lengua Española* (Vol. 1, pp. 695–706). Santiago de Compostela: Meubook.
- Castillo Lluch, M. (1996). *La posición del pronombre átono en la prosa hispánica medieval*. (Unpublished doctoral dissertation). Universidad Autónoma de Madrid, Madrid, Spain.
- Castillo Lluch, M. (2002). Distribución de las formas analíticas y sintéticas de futuro y condicional en español medieval. In M. T. Echenique Elizondo & J. P. Sánchez Méndez (Eds.), *Actas del V Congreso Internacional de Historia de la Lengua Española* (pp. 541–550). Madrid: Gredos.
- Eide, K. G. (2018). Postverbal subjects of unaccusative verbs in the history of Portuguese. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 149–171). Amsterdam: John Benjamins. (in this volume) doi:10.1075/ihtml.16.07eid
- Elvira, J. (1987). Enclisis pronominal y posición del verbo en español antiguo. *Epos*, 3, 63–79.
- Elvira, J. (1993). La función cohesiva de la posición inicial de frase en la prosa alfonsí. *Cahiers de linguistique hispanique médiévale*, 18/19, 243–278. doi:10.3406/cehm.1993.1089
- Fernández-Ordóñez, I. (2008–2009). Orden de palabras, tópicos y focos en la prosa alfonsí. *Alcanate*, 6, 139–172.
- Fischer, S. (2005). Construcciones con avance estilístico en el catalán antiguo. *Caplletra*, 38, 119–135.
- Fischer, S. (2010). *Word-order change as a source for grammaticalisation*. Amsterdam: John Benjamins. doi:10.1075/la.157

- Fischer, S. (2014). Revisiting Stylistic Fronting in Old Spanish. In A. Dufter & A. S. Octavio de Toledo y Huerta (Eds.), *Left sentence peripheries in Spanish. Diachronic, variationist and comparative perspectives* (pp. 53–76). Amsterdam: John Benjamins.
- Fischer, S., & Alexiadou, A. (2001). On Stylistic Fronting: Germanic vs. Romance. *Working Papers in Scandinavian Syntax*, 68, 117–145.
- Fontana, J. (1993). *Phrase structure and the syntax of clitics in the history of Spanish*. (Unpublished doctoral dissertation). University of Pennsylvania, Philadelphia, PA.
- Franco, I. (2009). *Verbs, subjects and Stylistic Fronting. A comparative analysis of the interaction of CP properties with verb movement and subject positions in Icelandic and Old Italian*. (Unpublished doctoral dissertation). University of Siena, Siena, Italy.
- Hrafnbjargarson, G. H. (2004a). *Oblique subjects and Stylistic Fronting in the history of Scandinavian and English: The role of IP Spec*. (Unpublished doctoral dissertation). University of Aarhus, Aarhus, Denmark.
- Hrafnbjargarson, G. H. (2004b). Stylistic Fronting. *Studia Linguistica*, 58(2), 88–134.
doi:10.1111/j.0039-3193.2004.00111.x
- Leonetti, M., & Escandell, M. V. (2009). Fronting and *verum focus* in Spanish. In A. Düfter & D. Jacob (Eds.), *Focus and background in Romance languages* (pp. 155–204). Amsterdam: John Benjamins. doi:10.1075/slcs.112.07leo
- Maling, J. (1980). Inversion in embedded clauses in Modern Icelandic. *Íslenskt mál og almenn málfræði*, 2, 175–193.
- Maling, J. (1990). Inversion in embedded clauses in Modern Icelandic. In J. Maling & A. Zaenen (Eds.), *Syntax & semantics: Modern Icelandic syntax* (pp. 191–215). London: Longman.
- Mathieu, E., & Sitaridou, I. (2002). Split *wh*-constructions in Classical and Modern Greek. *Linguistics in Potsdam*, 19, 143–182.
- Mathieu, E. (2006). Stylistic Fronting in Old French. *Probus*, 18, 219–266.
doi:10.1515/PROBUS.2006.008
- Menéndez Pidal, R. (1976a). *Orígenes del español. Estado lingüístico de la Península Ibérica hasta el siglo XI*. Madrid: Espasa-Calpe.
- Menéndez Pidal, R. (1976b). *Cantar de Mio Cid. Texto, gramática y vocabulario*. Madrid: Espasa-Calpe.
- Meyer-Lübke, W. (1900). *Grammaire des langues romanes. Vol. 3: Syntaxe*. Paris: H. Welter Éditeur.
- Molnár, V. (2010). Stylistic Fronting and discourse. *Tampa Papers in Linguistics*, 1, 30–61.
- Platzack, Ch. (1988). The emergence of a word order difference in Scandinavian subordinate clauses. *McGill Working Papers in Linguistics. Special Issue on Comparative Germanic Syntax*, 215–238.
- Rodríguez Molina, J. (2010). *La gramaticalización de los tiempos compuestos en español antiguo: Cinco cambios diacrónicos*. (Unpublished doctoral dissertation). Universidad Autónoma de Madrid, Madrid, Spain.
- Rohlf, G. (1949–1954). *Historische Grammatik der italienischen Sprache und ihrer Mundarten. I. Lautlehre. Band III: Syntax und Wortbildung*. Bern: A. Francke Verlag.
- Ślawomirsky, J. (1986). Wackernagel's Law in Old Spanish. *Revista Española de Lingüística*, 16(1), 91–100.
- Vennemann, T. (1974). Topics, subjects, and word order: From SXV to SVX via TVX. In J. Anderson & Ch. Jones (Eds.), *Historical linguistics: Proceedings of the First International Congress of Historical Linguistics, Edinburgh, September 1973* (pp. 339–376). Amsterdam: North-Holland.

Corpus

Admyte II. Archivo digital de manuscritos y textos españoles. Madrid: Micronet & Ministerio de Educación y Ciencia, 1999.

Alfonso, X. *Estoria de España* (I and II), Escorial Monasterio Y-I-2, Ll. A. Kasten & J. Nitti (Eds.), in *Admyte-II*.

Alfonso, X. *General Estoria* (II and IV), Madrid Nacional ms. 10237, W. L. Jonxis-Henkemans (Ed.), in *Admyte-II*.

Real Academia Española. *Corpus Diacrónico del Español (CORDE)*. Madrid: RAE. <<http://www.rae.es>>

The dative experiencer of Spanish *gustar*

Chantal Melis & Marcela Flores

Universidad Nacional Autónoma de México

This paper focuses on the dative experiencer of Spanish *gustar* ‘to please, like’, which replaced an older nominative experiencer. The syntactic shift calls for attention since it goes against the well-established diachronic tendency for oblique arguments to be eliminated in favor of nominative subjects (cf. English *like*). We find a partial explanation in the existence of a dative marking pattern on which *gustar* could model its behavior and try to identify the semantic and pragmatic factors that played a role in motivating the extension of the non-nominative experiencer to *gustar*. Our analysis confirms that lexical items retain traces of their source meaning which continue to shape subsequent developments, and that specific discourse contexts are instrumental in generating processes of change.¹

Keywords: Spanish; experiencer; dative; nominative; subject; oblique; quirky subject

1. Introduction

Languages from different families and areas have sentences that deviate from the canonical transitive model, in the sense that the entity perceived as subject appears with a dative case marking, while another element, in the nominative, controls verbal agreement. Consider these examples:

- (1) a. Russian:
Borisu nrvajatsja takie rubaški
BORIS.DAT like.3PL such shirts.NOM
‘Boris likes such shirts’ (Moore & Perlmutter, 2000)

1. We thank the two anonymous reviewers for their helpful suggestions and comments.

- b. Icelandic:
Henni leiddust strákar
 her.DAT bored boys.the.NOM
 ‘She found the boys boring’ (Sigurðsson, 2002)
- c. German:
Mir gefällt dieses Buch
 me.DAT pleases this book.NOM
 ‘I like this book’ (Haspelmath, 2001)

Sentences of this type are referred to in the literature as ‘dative’, ‘oblique’, ‘non-nominative’, or ‘quirky’ subject constructions. They have attracted the interest of linguists and have prompted a variety of analyses that seek to account for the non-nominative encoding of the subject-like participant. If no consensus has been reached as to the syntactic status of this participant, it is because the challenge posed by such sentences hinges on the more fundamental question of how one defines the notion of ‘subject’ (Fischer, 2004).

An important result that has emerged from the comparison between languages is that cross-linguistically these non-canonical ‘subject’ patterns tend to occur with similar event classes – mental experiences, psychological states, desire, need, possession, and existence – which share the property of involving an individual devoid of control over the verbal situation and in some sense affected by it (Onishi, 2001). Such a tendency clearly points to a phenomenon based on semantic principles.

At the same time, the underlying semantic principles prove somewhat elusive in all those cases where closely related verbs diverge in their case assignment patterns. The so-called psych verbs, as is known, provide good evidence on this topic, considering that the experiencer argument of these verbs is sometimes cast as a nominative subject, may appear in the form of a direct object, or takes dative (or oblique) marking in a construction likely to resemble those shown in (1) (see Haspelmath, 2001 for a typological overview of these possibilities). A familiar example of the syntactic variation associated with psych verbs is found in the contrast between English *like* (*I_{NOM} like beer*) and Spanish *gustar* (*Me_{DAT} gusta la cerveza* ‘I like beer’).

The present work brings Spanish *gustar* ‘to please, like’ back into the limelight in search of an explanation for its dative valence type. From a purely syntactic point of view, idiosyncratic case assignments are not to be ‘explained’; they simply require being stipulated in the verb’s lexical entry. On a different approach, however, one in which semantics and syntax are assumed to intersect, attempts at looking beyond the apparent randomness are justified. In the case of *gustar*, the quest for an explanation is sought in view of the fact that the verb shifted from a nominative to a dative experiencer pattern in the course of time, after passing through a

stage of alternating choices between options. Thus, in contrast to Modern Spanish, where *gustar* regularly appears in the dative pattern, as shown in (2):²

- (2) *me gusta la compañía de escritores y poetas*
 me.DAT like.3SG the company.NOM of writers and poets
 'I like the company of writers and poets' (1995, Adoum)

in former times, *gustar* took a nominative experiencer, as in (3):

- (3) *ella gustaba de su compañía*
 she.NOM liked.3SG of his company
 'She liked his company' (1631, Castillo Solórzano)

The evolution of Spanish *gustar* may be contrasted with the history of English *like*, which used to require a dative experiencer (*him liketh*) and later converted to a nominative subject verb (*I like*) (Allen, 1995). The path taken by English *like* conforms to the general diachronic tendency observed in the European languages, according to which dative or oblique experiencers gradually came to be replaced by nominative subjects, defined as more 'regular' from the perspective of a transitive language system (Cole et al., 1980; Haspelmath, 2001; Seefrantz-Montag, 1984).

Movements in the opposite direction, as exemplified by *gustar*, are known to have occurred at earlier stages of Germanic (Allen, 1995; Bardđal, 2009) and Romance history (Bauer, 2000), but have received less attention in the literature. Scholars attribute them to the attraction of regular (nominative) verbs by existing lexical units associated with a non-canonical marking pattern. The explanation can be extended to Spanish *gustar*, which, as we shall see, likewise modeled its behavior on the older dative experiencer verb *placer* 'to please, like'.

The syntactic shift undergone by *gustar* does not, in this sense, introduce a 'new' pattern into the language. In fact, Spanish possesses a series of verbs behaving like *gustar*, some of which have been around since early times, while others have joined the dative group at later stages (Melis & Flores, 2013). But the existence of a constructional model in itself does not explain why specific items are attracted to it. It does not account for the forces at play in triggering the change of

2. All Spanish examples cited in this paper are extracted from the database of the *Real Academia Española*. The *Corpus de referencia del español actual* (CREA) is the source of the 20th century citations, whereas the *Corpus diacrónico del español* (CORDE) has provided the examples corresponding to earlier centuries. In order to carry out the analyses that will be presented in the following sections, we gathered representative samples of the verbal uses along the diachronic axis. We made sure that our samples included the different forms (in terms of person, number, tense, and mood) of the respective verbal paradigms, and gave preference to texts written in prose.

grammatical behavior. In her diachronic work on the psych verbs of English, Allen (1995) wonders about the diverging pathways of *like* (*I like it*) and *please* (*it pleases me*), which she proposes to elucidate by delving into the respective meanings of the verbal items and their contexts of use. Our aim is to carry out a similar study focusing on Spanish *gustar* in order to gain insight into the motivations behind its syntactic shift.

We will examine the semantics of *gustar*, with special attention to the original meaning of its Latin source ('taste'), under the assumption that lexical items retain some traces of earlier senses which continue to shape their subsequent developments (Bybee & Pagliuca, 1987; Hopper, 1991). We will be equally interested in the discourse environments which accepted the use of the dative pattern in the initial phase of the evolutionary process, given the hypothesis that lexical units undergo change within highly specific contexts and constructions (Bybee et al., 1994; Diewald, 2002; Heine, 2002). And we will glance at two other closely related verbs of 'liking', namely *amar* 'to love' and *querer* 'to want, like, love', whose stable association with a nominative subject experiencer will be contrasted with the realignment of *gustar* so as to deepen our understanding of why *gustar* alone yielded to the pressure exerted by *placer*.

The paper is organised as follows. Section 2 provides an overview of the history of *gustar* and a first approach to its syntactic shift. In Section 3 we compare the distinct trajectories of the Spanish 'liking' verbs. Section 4 contains the analysis of the semantic and pragmatic factors that motivated the change undergone by *gustar*. In Section 5 we add a quick note regarding the 'love' semantics of the verbs under study, and in Section 6 we present our conclusions.

2. The historical development of *gustar*

The primary goal of this paper, as we mentioned in the Introduction, is to explain the morphosyntactic shift experienced by Spanish *gustar* 'to please, like' from a semantically oriented perspective. Before taking up the semantic analysis, we will give an overview of the history of *gustar*, which actually comprises three major stages: (1) *gustar* begins as a 'taste' verb, semantically; (2) joins the subclass of the Spanish 'liking' verbs, associated with a nominative experiencer (like *amar* 'to love' and *querer* 'to want, like, love'); and (3) adopts the dative construction of *placer* 'to please, like'. This evolution is the topic of the present section.

2.1 The physical sense of taste

Spanish *gustar* descends from the Latin verb GUSTĀRE meaning 'to taste, to take a little of'. A special point of interest lies in the fact that the Latin source of *gustar*

was occasionally extended to non-physical contexts, as in (4), where it conveyed a more abstract idea akin to ‘taste, partake of, enjoy’ (Lewis, 1889, s.v. *gustō*):

- (4) *paululum istarum artium gustavi*
 a little these theories.GEN tasted.1SG
 ‘I tasted a little of these theoretical studies’ (Cic. *Or.* 3.75)

The metaphorical extension comes as no surprise. In her analysis of the Mind-as-Body Metaphor, Sweetser (1990, p. 37) remarks that taste “is a physical sense which seems universally to be linked to personal likes and dislikes in the mental world”. According to Sweetser, the explanation for this universal link-up must be sought in the properties of closeness (contact with the sensed object) and subjectivity (tastes vary across people), associated with the sense of taste, which facilitated the mapping of the physical domain onto the intimate and personal realm of the emotions.³

Turning to Spanish, we find that throughout the Medieval Period of the language (11th to 15th century) *gustar* is used primarily as a verb within the semantic field of ‘tasting’ or ‘eating’ (Melis, 1998):

- (5) a. *E non gosto todo el pueblo pan fasta*
 and not tasted.3SG all the people.NOM.SG bread.ACC until
la noche.
 the night
 ‘And nobody ate bread until the evening’ (ca. 1400, *Biblia ladinada I-i-3*)
- b. *non puede todo catar nin gostar de quanto les dava*
 not can.3SG all sample nor taste.INF of as much them gave
a comer
 to eat
 ‘He could not sample nor taste everything he was given to eat’
 (ca. 1253, *Sendebär*)

As the examples in (5) show, the tasting agent is given the grammatical role of (nominative) subject, whereas the thing sensed commonly functions as the direct object of the transitive clause (5a), but may be introduced by the preposition *de* when a partitive reading is intended (5b).

In line with its Latin source, Spanish *gustar* is sometimes extended to more abstract spheres of meaning (Vázquez Rozas & Rivas, 2007). Depending on the

3. Elaborating on Sweetser’s proposal, Ibarretxe-Antuñano (1999, p. 82) further observes that the sense of taste implies a process of “fine discrimination” between a whole range of slightly different (mixtures of) flavours, suitable to the purpose of describing the variable likes and dislikes of human beings in other domains.

context, the metaphorical uses evoke a notion of ‘experiencing’ something, or may acquire definite overtones of ‘pleasure’ (cf. Ibarretxe-Antuñano, 1999), as illustrated in (6a) and (6b), respectively:

- (6) a. *Aristotiles, Juuenal, e otros muchos que estudiaron [...]*
 Aristoteles Juvenal and others many who studied
este vaso de amargura cruelmente lo gustaron
 this cup of bitterness cruelly it.ACC tasted.3PL
 ‘Aristoteles, Juvenal, and many other scholars had a taste of this bitter cup’ (1424, Cancionero de Juan Fernández de Íxar)
- b. *non podríamos conoçer ni menos gustar la dulçesa*
 not could.1PL know nor less taste.INF the sweetness.ACC
de la misericordia de Dios
 of the mercy of God
 ‘We could not know, let alone taste, the sweetness of God’s mercy’ (1450, Cartagena)

2.2 Extension to the mental world

Our corpus data indicate that *gustar* becomes established as a psychological verb, meaning ‘to enjoy, take pleasure, like’, in the second half of the 16th century. The new meaning has obviously been prepared by the metaphorical examples shown in (6), but the integration of the new meaning into the semantic structure of the verb is not made clear until then, when the mental use of *gustar* undergoes a sharp increase in frequency relative to the physical sense, and grows more diverse. Some examples are given in (7):

- (7) a. *como los yndios no saven leer gustan más de la*
 since the indigenous.NOM not know read like.3PL more of the
pintura que de la escriptura
 painting than of the writing
 ‘Since the indigenous can’t read, they like painting more than writing’ (1575, Moya y Contreras)
- b. *¿por qué gustas de tratar-me d-este modo?*
 why like.2SG of treat-me of-this way
 ‘Why do you like to treat me this way?’ (1615, Cervantes)
- c. *Dicen que el Príncipe gustó mucho de la fiesta*
 say.3PL that the Prince.NOM liked.3SG much of the feast
 ‘They say that the prince enjoyed the party very much’ (1634, Vilches)
- d. *las que gustan de que las galanteen los hombres*
 REL.3PL.FEM like.3PL of that them court the men
 ‘[The ladies] who like to be courted by men’ (1657, Zabaleta)

As we see in (7), the meaning extension has not altered the morphosyntactic behavior of *gustar*, which continues to assign the subject function to the salient human participant, the former tasting agent reinterpreted as the experiencer of the mental state.

The second participant now instantiates the ‘stimulus’ of the mental experience (Talmy, 1985). The stimulus may refer to a (nominal) thing (cf. *pintura*, *escriptura* (7a) and *fiesta* (7c)), or may evoke a situation, specified in an infinitive (cf. *tratarme* (7b)) or a tensed clause (cf. *que...* (7d)). Note that the stimulus is almost always introduced by the preposition *de*, whereas the corresponding argument of physical *gustar* tended to function as direct object (see Vázquez Rozas & Rivas, 2007 for relevant distributional data). The difference can be attributed to the fact that the second participant has shifted from being an affected patient – a substance taken into the mouth and ingested – to fulfilling the role of a mental stimulus, dissociated from any notion of change (Dowty, 1991). The reduced transitivity in the semantics of the event is what the use of the preposition nicely reflects on the syntactic level.

One final aspect worthy of mention is that mental *gustar* selects for a non-human stimulus in the overwhelming majority of its uses. The combination with a human referent does occur, as in (8), but scarcely. This fact will be of crucial importance for the subsequent evolution of *gustar*:

- (8) *todo gran hombre gusta de los grandes hombres*
 every great man.NOM like.3SG of the great men
 ‘Every great man likes great men’ (1657, Gracián)

2.3 The syntactic shift

The textual evidence leads us to situate the emergence of the dative experiencer pattern with *gustar* towards the end of the 18th century. It is true that a few isolated uses of the pattern are attested in earlier centuries, and one way to approach these exceptional occurrences is thinking of them as “exploratory expressions”, susceptible of causing a process of change if they are replicated and spread (Harris & Campbell, 1995).

In the case of *gustar*, the signals of a change in process correspond to the late 18th century, when the dative form enters in competition with the older nominative experiencer, as shown in (9):

- (9) a. *ya que gustas de coplas*
 since like.2SG of verses
 ‘Since you like verses’ (1779, *Vexamen en pie de romance*)
 b. *¡Me gustan tanto los versos!*
 me.DAT like.3PL so much the verses.NOM
 ‘I like verses so much!’ (1792, Fernández de Moratín)

In the first phase of the development, although the presence of the dative is clearly felt, the nominative experiencer continues to function as the overwhelmingly preferred option.⁴ With the progress of time, however, the dative experiencer becomes increasingly more frequent, at the expense of the nominative subject of *gustar*.⁵ The outcome of the gradual replacement of the old construction with the new case marking pattern is the situation reflected by contemporary Spanish, where the dative experiencer appears in more than 95% of the occurrences of *gustar*, the original nominative experiencer being confined to a residual sphere of use (see below Section 4.2).⁶ In sum, compared to the well-known evolution of the English verb *like* (*him liketh* → *he likes*), Spanish *gustar* projects a movement that went in the opposite direction.

As we said earlier, the reverse movement of *gustar* has to be understood in light of the existence in Spanish of a non-canonical marking pattern (a type of ‘dative subject’ construction), which was initially associated with a few verbal items and through time spread to an ever-increasing number of predicates (Melis & Flores, 2013). In the particular case of *gustar*, the influence came from dative experiencer *placer* ‘to please, like’ (Melis, 1998; Vázquez Rozas & Rivas, 2007). *Placer* can be traced back to the oldest stage of the Spanish language, and is found with great frequency in the medieval texts (Melis, 1999). The experiencer argument of *placer* bears dative case in all its occurrences, whereas the stimulus is typically introduced by a preposition at first (10a), but is often nominative later on (10b):⁷

- (10) a. *a los angeles plaze mucho con la su virginidad*
 DAT the angels please.3SG much with the her virginity
 ‘The angels are very pleased with her virginity’ (1293, *Castigos*)
- b. *a otros plazen las guerras*
 DAT others please.3PL the wars.NOM
 ‘Others like war’ (1436, *Cancionero de París*)

4. In our data corresponding to the final decades of the 18th century, the occurrences of the dative pattern make up 25% of the registered uses of *gustar* (106 of a total of 423 examples).

5. By the end of the 19th century, the occurrences of *gustar* with dative experiencer have risen to 64% (312 of a total of 484 examples).

6. Vázquez Rozas and Rivas (2007) report 97.44% in their database; these numbers coincide with ours (399 of a total of 412 examples = 96.84%).

7. Spanish *placer* descends from Latin PLACĒRE ‘to please’, a semi-impersonal verb governing a dative experiencer. The valence frame of *placer* in Spanish (EXP/DAT + STIM/PP) is modeled on the impersonals of Latin, which, like *piget* (‘it irks’) or *puDET* (‘it makes ashamed’), were constructed with a non-nominative (accusative) experiencer and a genitive stimulus (Melis & Flores, 2013).

Towards the end of the Middle Ages (late 15th century), however, *placer* begins to decline, both in terms of frequency of use, as in relation to its distributional patterns (see below note 11). The culmination of this process is reached by the end of the 17th century, when *placer* appears to have practically fallen into disuse (Melis, 2011).⁸

It is important to note here that the decline of *placer* coincides with the period in which *gustar* becomes established as a verb of liking in the mental lexicon of Spanish. We will return to this point in the next section.

3. Construals of mental events

As background to our semantic analysis, it will be helpful to recall that mental verbs denote complex events, which come about as a result of the interaction between the experiencer, who directs attention to the stimulus, and the stimulus, which, in turn, induces an emotional reaction or cognitive judgment in the experiencer (Croft, 1991; Dowty, 1991; Kemmer, 1993; Langacker, 1991; Maldonado, 1999). This two-sided relation, characteristic of mental experiences, is what makes such experiences susceptible of being construed in different ways. That is to say, mental events can be couched into distinct morphosyntactic frames which impose a particular perspective on the same reality; different frames have the effect of foregrounding some components of the mental scene and leaving some other ones out of perspective.

The contrast relevant to our analysis hinges on which of the two verbal participants is presented as more responsible for the experiencer's internal state. Thus, with the experiencer as the grammatical subject, because of the typical association of this function with agency, the construction highlights the active involvement of the human participant, at the expense of the stimulus. But if the experiencer is coded dative, given the symbolic value of this case marker (Dąbrowska, 1997; Næss, 2007), the emphasis shifts to the 'affected' side of the human participant, with a concomitant allusion to the role played by the stimulus in engendering the mental event within the experiencer (cf. Calderón Campos, in this volume, on the use of the definite article as a textual prominence marker when used with proper names).

This said, we now propose having a quick look at two other members of the Spanish subclass of 'liking' verbs, *amar* 'to love' and *querer* 'to want, like, love',

8. *Placer* has not totally disappeared from the language; it still shows up from time to time in written texts.

which, unlike *gustar*, were not attracted by *placer*. A close examination of these quasi-synonyms of *gustar* will reveal subtle differences in meaning that will help us account for their distinct histories. After examining *amar* and *querer*, opposed to *gustar*, we revert to *placer* in order to stress the affinities *gustar* shared with the dative experiencer verb with which it was drawn in line.

3.1 *Gustar* vs. *amar*

Spanish *amar* ‘to love’ (< Lat. AMĀRE ‘to love’) is typically found in constructions with a human stimulus serving to express the stable and deep relationships which are thought to define parental, marital, or divine love. The early examples given in (11) illustrate the type of contexts to which *amar* will tend to be restricted during the entire course of Spanish history:

- (11) a. *assi cuemo Christo nos amo*
 so as Christ.NOM us.ACC loved.3SG
 ‘In the same way that Christ loved us’ (1260, *El Nuevo Testamento*)
- b. *e que-s amassen como madre e fijo*
 and that-REFL love.SUBJ.3PL like mother and son
 ‘And that they’d love each other like mother and son’ (1275, Alfonso X)
- c. *d-alli adelant amo a su mugier donna Helisabet*
 from-there on loved.3SG ACC his wife lady Elizabeth
muy mas que non solie
 very more than not used
 ‘from then on he loved his wife Elizabeth much more than he used to’
 (1270, Alfonso X)

Nothing predisposed *amar* to such specialisation – the French cognate *aimer*, for example, shows a much wider distribution of uses, but the constraints imposed on the verb in the Spanish language are crucial to an understanding of why the transitive template EXP/S – STIM/DO, inherited from Latin, was never exposed to any change.⁹ We have to realise, indeed, that Spanish *amar* focuses on tight loving bonds, which imply a notion of unity, favorable to the endurance of the

9. French *aimer* is used with non-human referents, nominal (i) or clausal (ii), to denote mental experiences similar to those expressed by English *like* or Spanish *gustar*:

- (i) *Il aime aussi les bonbons.*
 ‘He also likes sweets’
- (ii) *Il n’ aime pas (à) sortir seul.*
 ‘He doesn’t like going out alone’

relationship (Kövecses, 2000), in conjunction with a sense of commitment, feelings of care and concern, and the desire to do good things for the person one holds dear (Wierzbicka, 1992). *Amar* thus evokes the love people feel, independently of the properties or actions of the loved one, and, in accordance with this view, the verb has the accent put on the sensation of the experiencer, at the expense of the human stimulus, whose participation in creating and sustaining the affective bond is maximally backgrounded. Hence, the reason why *amar* was not affected by *placer* is because the transitive frame, within which the subject experiencer is construed as directing his/her internal feelings towards the patient-like object of love, suited Spanish *amar* exceptionally well.

Gustar, on the other hand, never came to express this kind of love, given its preference for inanimate stimuli. The inanimate choice makes sense in light of the original meaning of *gustar* – human beings are not ‘tasted’ nor ‘savored’, and had a profound influence on the evolution of *gustar*, as we will see below. For the moment, it suffices to say that pleasurable experiences involving things or situations tend not only to be more transitory than our feelings for people, but also more dependent on specific properties of the stimulus. Hence, a construal bestowing greater salience on the stimulus may come to be judged as more adequate for expressing these experiences than one in which the stimulus figures as the passive target of a person’s feelings.

3.2 *Gustar* vs. *querer*

Turning to Spanish *querer* ‘to want, like, love’, it is important to mention that we are here dealing with a basic verb of ‘wanting’. In principle, ‘wanting’ verbs and ‘liking’ verbs belong to different classes. According to Dixon (1991), the latter (*like, love, hate, etc.*) designate a certain feeling an experiencer gets about a stimulus (Dixon, 1991, p. 155), whereas the former (*want, wish (for), desire, etc.*) express an attitude towards some event or state “that is not (yet) real” (Dixon, 1991, p. 184).

As a ‘wanting’ verb, *querer* is normally construed with a complement clause that specifies the desired, i.e. the not (yet) real, event or state:

- (12) a. *queremos que lo sepan todos*
 want.1PL that it know all
 ‘We want everyone to know it’ (1290, *Fuero Burgos*)

On occasion, Spanish *amar* too appears with a non-human stimulus in slightly hyperbolic utterances; the object of love always refers to grand or noble entities: cf. *justicia & verdat* ‘justice and truth’ (CORDE, 1260), *los libros* ‘books’ (CORDE, 1663), *a mi nación* ‘my country’ (CREA, 1997).

- b. *le replicó el señor y dixo: si quieres ser perfecto,*
 him answered the lord and said if want.2SG be.INF perfect
anda, vende quanto tienes
 come sell as much have

‘The Lord responded to him saying: if you want to be perfect come, and sell what you have’ (1650, Solís y Valenzuela)

However, even if the two concepts need to be kept separate, it is clear that the boundary between ‘wanting’ and ‘liking’ is not tight. The interconnection between the two semantic domains (cf. Bruce & Bruce, 2010) derives from the fact that our positive feelings of happiness, pleasure or love have much to do with the fulfillment of our wishes; we feel good when things happen that we wanted to happen, or when we obtain or possess the object we desired (cf. Wierzbicka, 1992). The dictionaries reflect this conceptual proximity by subsuming the senses ‘to desire’ and ‘to wish’ under the entry of the English verb *like*.

Spanish displays a similar kind of conflation with both *querer* and *gustar*: *querer* developed a ‘liking’ sense, restricted to the contexts where the verb selects a human stimulus (see below Section 5), whereas *gustar*, once established as a mental verb, expanded its ‘liking’ sense to the domain of desire, in constructions in which the stimulus designates an event or state “that is not (yet) real”:¹⁰

- (13) *Si quiere saber quién soy, dama de los ojos negros escúche-me*
 if want know who be.1SG lady of the eyes black listen-me
y lo sabrá, si es que gusta de saber-lo
 and it will.know if is that like.2SG of know-it

‘If you want to know who I am, lady with the black eyes, listen to me and you will know, if indeed you like/wish to know’ (1580, Góngora y Argote)

In these contexts *gustar* competed with *querer*. A careful analysis of the alternation between the two verbs would probably reveal that the speaker who chose *gustar* over *querer* sought to accentuate the experiencer’s feeling of delight at the thought of seeing the event come true.

For our purposes, the significant factor is that *querer*, in contrast to *gustar*, never showed the slightest tendency towards developing a dative experiencer. In itself, the basic ‘wanting’ sense of *querer* cannot be held responsible for the

10. When *querer* combines with a noun phrase of inanimate reference, the object of desire shares the *irrealis* status of the propositional stimulus illustrated in (12), that is, it refers to an object that the experiencer wants to ‘get’, ‘receive’ or ‘have’. This is typical of verbs of ‘wanting’ (Dixon, 1991, p. 92). Cf. *No quiero tu agradecimiento, sino tu amor* ‘I don’t want your gratitude, but your love’ (CREA, 1999).

divergence in evolutionary pathways, considering that, typologically, verbs of wanting are good attractors of non-canonical case marking patterns (Onishi, 2001). In fact, Spanish *placer*, with its dative experiencer, was used with a certain amount of frequency to render a ‘wanting’ notion.¹¹

So we must turn to the etymology of *querer* to find an explanation. With respect to the verb’s origin, Corominas & Pascual (1981, s.v. *querer*) state that *querer* descends from the agentive Latin verb *QUAERERE* ‘to seek, to look for, to ask for’, which in Spanish and Portuguese came to mean ‘want’, in contrast to the other Romance languages, where the same notion continued to be indicated by the heirs of Latin *VELLE* ‘to will, wish, want’. Taking this into account, it can then plausibly be argued that the feature of agentivity retained from the Latin source played a role in maintaining the ‘wanting’ verb closely tied to the agent-like subject experiencer of the transitive construction.

3.3 *Gustar* and *placer*

Having contrasted *amar* and *querer* with *gustar*, let us briefly go back to *placer* in order to reflect on the existing parallels with *gustar* and their role in the process of change under discussion.

Obviously, one important feature *placer* and *gustar* share in common is their nearly synonymous meaning focusing on a pleasurable experience. In addition, both *placer* and *gustar* have a definite preference for the stimulus of inanimate reference (things or situations) and rarely take a human entity as their second participant. Furthermore, as we mentioned, the two verbs have been habilitated to

11. With *placer*, the ‘wanting’ reading typically emerges in contexts where the verb focuses on the experiencer’s reaction to a certain demand or proposal, and communicates the willingness on the part of the subject to go along with the suggested course of action:

- (i) *E la donzella le dixo que le plazía de buena voluntad de le responder, e dixo luego luego sin detardar al sabio [...]*

‘And the maiden told him that she was very pleased/willing/eager to give him an answer, and at once without delay she said to the scholar [...]

(ca.1250, *apud* Melis, 1999)

It is worth pointing out that the process of decline of *placer*, which we mentioned in Section 2.3, was characterised by an increase in frequency of the ‘wanting’ sense, to the detriment of the verb’s ‘liking’ sense. This increase went hand in hand with the progressive reduction of the ‘wanting’ sense to the concept of the will of God, evoked in formulaic expressions (cf. *plega a Dios que [...]* ‘would to God that [...]; *placiendo a Dios* ‘God willing’, etc.). Shortly before *placer* falls into disuse, the majority of the documented examples consist of such formulas (Melis, 2011).

convey a ‘wanting’ notion in certain types of contexts. Finally, there is the coincidence in time between *placer*’s process of decline and the period in which *gustar* becomes established as a new verb of liking in the mental lexicon. In view of these convergences, it is tempting to suggest that *gustar* was drawn to the emotional domain to compensate for the loss of the older medieval verb. Regardless of this hypothesis, the interconnections between the two verbs shed light on the circumstances that enabled and facilitated the syntactic shift of *gustar* towards the dative experiencer construction.

4. The change and its motivation

In the present section we examine how the substitution of the nominative subject pattern by the dative construction was implemented. Since we are interested in the semantic motivation behind the syntactic change, we have to look closely at the contexts in which the dative construction begins to appear. These contexts are likely to exhibit some particular features whose compatibility with the innovative pattern must be felt as unique or exceptional. Something along these lines has to be assumed since speakers, in a more or less conscious way, are giving preference to the dative construction over the older nominative experiencer pattern in those specific environments. The alternation between uses further implies that the two syntactic templates, now in competition, are associated with a specific semantics responsible for the distinct options speakers resort to depending on the context.

From another perspective, we also have to wonder about the ways in which the original nominative experiencer verb was made to adjust to the semantics of the new syntactic frame. The older *gustar* had its focus placed on the experiencer who takes pleasure in something; within the new valence type the experiencer is construed as ‘affected’ by the impact of a stimulus. This hardly defines a minor change in viewpoint. So there will be need to probe into the semantic structure of *gustar*, searching for the relevant meaning features that explain the ease with which the ‘liking’ verb shifted from one focus to another.

All these issues will be addressed in Section 4.1. The gradual spread of the dative experiencer, to the detriment of the nominative subject, will be outlined in Section 4.2.

4.1 The initial contexts

The analysis of the contexts that initially attract the dative construction modeled on *placer* reveals one major tendency: the new pattern surfaces where the stimulus participant of *gustar* has topical status. In some cases, the stimulus corresponds to

an authentic *discourse* topic; it acts as the pivotal entity around which a more or less extended fragment of the discourse is organised. Such occurrences, however, constitute the minority of the examples analysed, given that inanimate referents, as is well known, rarely manage to stay in the foreground for a substantial length of time. Typically, the stimulus functions as a more ephemeral *clausal* topic, enjoying the prominence of the element that the sentence is ‘about’.

The following examples will serve us to illustrate the topicality of the stimulus. In (14a) the *little work* occupies center stage within a whole textual unit. (14b) and (14c) display instances of short-lived focuses of attention, whose referents are either anaphorically or deictically anchored in the discourse. These constitute the typical group of documented cases. Finally, (14d) is meant to show that our category of topics includes lexically coded entities which are part of the semantic frame evoked by a previously mentioned element:

- (14) a. *He visto una obrita en dos tomos[...] me ha*
 have.1SG seen a little.work in two volumes me.DAT has
gustado infinito [...] luego se la envío á vd., porque
 pleased infinitely later CL.DAT it send.1SG DAT you because
sé que le gustará
 know.1SG that you.DAT will.please.3SG
 ‘I have seen a little work in two volumes [...], I have found it extremely
 pleasing [...], I’ll send it to you later, because I know it will please you’
 (1768, Azara)
- b. *Rindo a vuestra merced mil gracias por la carta de*
 render.1SG to your mercy thousand thanks for the letter of
Palafox, que me ha gustado mucho
 Palafox REL me.DAT has pleased much
 ‘A thousand thanks to your grace for the letter of Palafox, which really
 pleased me’
 (1759, Cunha)
- c. *Prosigue amigueta mía; que me gustan*
 proceed little.friend mine that me.DAT please.3PL
esas reglas
 these rules.NOM
 ‘Go on, my dear friend; I like these rules’
 (1788, Iriarte)
- d. *y sólo de Vieyra he leído algunos sermones, porque*
 and only of Vieyra have.1SG read some sermons because
me gustan mucho sus agudezas
 me.DAT please.3PL much their witticisms.NOM
 ‘And I’ve only read some of Vieyra’s sermons, because I like their bite’
 (1758, Isla)

What the examples in (14) have in common is that the stimulus is given the highest degree of prominence in the clause built with *gustar*. The importance of the stimulus is enhanced by the fact that the dative construction ascribes subject features to this participant (the stimulus is nominative and controls verbal agreement). Beyond the privileged treatment granted to the stimulus, these sentences furthermore invite an inference to the effect that the stimulus possesses certain characteristics which are to be seen as responsible for the sensation of pleasure registered by the experiencer. That is to say, what is becoming clear is that the mental state denoted by *gustar* is no longer viewed from the perspective of the experiencer who ‘takes pleasure’ in something, but is perceived instead from the vantage point of the stimulus that ‘causes pleasure’ to the experiencer. This follows naturally from the emphasis on the topical stimulus. At the same time, the shift in point of view sheds considerable light on the motivation for the use of the new dative pattern in these specific contexts, where the older construction, with its focus on the agent-like experiencer, would seem far less fitted.

Significantly, as a result of its insertion into this type of context, *gustar* acquires a new shade of meaning, whereby the notion ‘be pleasing, please’ comes to join the older sense ‘find pleasant, like’. This constitutes a change in the semantic structure of the verb, one that from then on will enable *gustar* to accommodate alternations in viewpoint determined by the choices of speakers. In this way, the evolution of *gustar* lends support to the hypothesis that change does not affect individual lexemes, but rather lexemes embedded in specific contexts (Bybee et al., 1994; Diewald, 2002; Heine, 2002).

The other question we should address relates to the surprising ease with which *gustar* was able to adjust to the shift in perspective. Under our proposal, the explanation emerges from the semantic domain of the verb’s original meaning. In the physical experience of taste, indeed, the sensation felt by taking something into the mouth cannot be dissociated from the substance whose sweet, sour, salty, or bitter quality produces the sensation. Languages reflect this by having one word, like English *taste*, denoting both the sensation of the tasting person and the flavor of the tasted object. Mapped onto the mental world, this particular feature implies that the notion of ‘liking’ will be able to refer to the experiencer who finds something pleasant, as well as to the stimulus whose special properties cause the appeal. The fact that mental *gustar* had no difficulty in moving from one reference to the other suggests that this particular feature was ‘retained’ (Bybee & Pagliuca, 1987; Hopper, 1991) from its original sense.

Table 1 summarises the results of our analysis of the stimulus in the first dative contexts, associated with the final decades of the 18th century.¹² The distribution

12. In footnote 4 above we mentioned that the dative pattern was found in 106 registered examples (as opposed to 317 examples with a nominative experiencer). For the detailed analysis

of special interest concerns the inanimate ‘things’, which we have divided into topical (83%) and non-topical (17%) entities. At this initial stage, the occurrences of clausal stimuli referring to ‘situations’ are exceptionally scarce (9%), due to the unlikelihood of this referential category to acquire topic status. The low numbers for the human stimulus (14%), on the other hand, are consistent with *gustar*’s preference for the non-human kind of stimulus throughout the history of Spanish.

Table 1. Topicality of stimulus in initial dative contexts (late 18th century)

	#		% of total
Human	36		14
Thing	191		77
		% of ‘things’	
[+ TOP]	158	83	
[- TOP]	33	17	
Situation	22		9

Total of occurrences of stimulus: 249

The initial contexts bring into view a second tendency, which has to do with the referent of the experiencer argument: speakers are more likely to select the dative pattern over the older construction when communicating something about their own mental state. The examples cited in (14) gave a good sense of this preference. To account for it, we first need to recall that the dative case entails a notion of affectedness for the referent cast in that role. Actually, all experiencers, as the loci of internal processes, are to be thought of as affected in some sense; yet, contrary to the nominative case, the dative signals this feature overtly. Next, consider the moment in which the dative is just beginning to compete with the nominative experiencer. Such a scenario implies that speakers make choices, and are therefore particularly sensitive to the meaning features of the new syntactic template. Finally, it is important to stress that exposing a condition of affectedness presupposes an ‘internal’ view, that is, the view of someone who has direct access to the experiencer’s self and is aware of the pleasing effect the stimulus has on the experiencer, because of his or her ‘empathy’ (Kuno & Kaburaki, 1977) with the referent.¹³

of the first contexts of use of the dative, however, we considered it necessary to have a larger sample. The sum of (exclusively) dative tokens was raised to 249, through a new search in CORDE guided by the criteria exposed in footnote 2. The data summarised in Table 2 belong to the same sample as those of Table 1.

13. Interestingly, Sadler (2007) observes a similar bias towards first person referents in the development of the dative experiencer construction in Japanese, for which the author proposes an explanation analogous to what we are suggesting here.

Gathering all these facts, the concentration of first person experiencers within the dative use can be explained.

In Table 2 we show the distribution of the referents of the dative argument. The data confirm that the new dative pattern favors first person experiencers.¹⁴ They also reveal that, in the first phase of the development of the dative construction, the number of clauses that lack an explicit reference to the experiencer is surprisingly high (13%). These results tie in with the overwhelming presence of the topical stimulus, which on more than one occasion functions as a unique focal point, to the exclusion of the experiencer, understood to refer to the whole of humanity.

Table 2. Speaker empathy with experiencer in initial dative contexts (late 18th century)

	#	% of total
1 p. Exp.	123	49
2 p. Exp.	38	15
3 p. Exp.	56	22
(implicit)	32	13

Total of occurrences of stimulus: 249

4.2 Generalisation of the dative construction with *gustar*

As we saw in Section 2.3, the dative pattern will eventually displace the original nominative experiencer construction and establish itself as the regular valence frame of *gustar*. Between the phase of emergence (late 18th century), and the moment at which the dative experiencer converts into the norm (mid-20th century), a process of gradual expansion takes place, in which the dative variant spreads to an ever-widening range of contexts. In correlation with this increase in productivity, the new construction becomes more and more dissociated from the semantic and pragmatic meanings that triggered its development at the initial stage, and, by the same token, is brought each time nearer to the use of a conventionalised grammatical structure.

14. The skewed distribution can be appreciated even better if we compare the results obtained for *gustar* with nominative experiencer in the same period: 14 occurrences (10%) of first person subjects in a total of 138 examples of *gustar* used in its 'liking' sense (the 'wanting' examples have been left out because the new dative pattern does not convey this notion; see below).

If we look at the data from the end of the 19th century, we are able to observe an on-going movement that is pushing in that direction.¹⁵ We notice, for example, that the dative construction no longer privileges the topical stimulus, and that, as a result of its expansion to less prominent entities, it is now used more freely with ‘situations’ coded in the form of a clause (see Table 3).

Table 3. Extension to non-topical stimulus in late 19th century

	#		% of total
Human	55		18
Thing	179		57
		% of ‘things’	
[+ TOP]	55	31	
[- TOP]	124	69	
Situation	78		25

Total of occurrences of stimulus: 312

Turning to the experiencer, we likewise perceive that the affectedness feature associated with the dative case has lost something of its specific informative value, as reflected by the fact that the dative use is expanding to third person referents, whose internal state is not directly accessible to the speaker (see Table 4).

Table 4. Extension to third person experiencer in late 19th century

	#	% of total
1 p. Exp.	120	38
2 p. Exp.	50	16
3 p. Exp.	128	41
(implicit)	14	4

Total of occurrences of experiencer: 312

Since the progression of the dative pattern is happening at the expense of the older construction, we are also interested in finding out how the use of *gustar* with nominative experiencer is being affected. What we discover here is that the process

15. The corpus of 312 dative examples from the CORDE, representative of the late 19th century, is identical to the one mentioned in footnote 5.

of change is generating some kind of split, according to which the expression of the ‘liking’ sense is being taken over by *gustar* with dative experiencer, while the original construction remains in charge of the semantic domains which are not in dispute, that is to say, the physical sense of taste, as in (15a), the metaphors tied to the source meaning, as in (15b), and, most importantly, the ‘wanting’ sense, as in (15c), which *gustar* with dative experiencer will never appropriate:¹⁶

- (15) a. *Hacía tiempo que no gustaba un café tan exquisito*
 ago time that not tasted.1SG a coffee.ACC so exquisite
 ‘I hadn’t tasted such an exquisite coffee in a long time’ (1889, Palma)
- b. *oyendo y gustando con fruición inmensa la*
 listening and tasting with fruition immense the
deliciosa música
 delicious music.ACC
 ‘Listening to, and enjoying with great pleasure the delightful music’
 (1885, Pérez Galdós)
- c. *Si te incomodo, me lo dices; si gustas de estar sola,*
 if you bother.1SG me it say.2SG if like.2SG of be.INF alone
me voy á mi cuarto
 REFL go.1SG to my room
 ‘If I bother you, tell me; if you like/wish to be alone, I’ll go to my room’
 (1892, Pérez Galdós)

With the progress of time, the nominative experiencer construction will continue to lose ground, suffering a series of losses that will reduce its sphere of use dramatically. The consequences of this development are clear to see in contemporary Spanish, where *gustar* is hardly ever employed to refer to the physical sense of taste, *probar* ‘to taste, try, sample’ being chosen instead (Ibarretxe-Antuñano, 1999), and where *gustar* no longer competes with *querer* to render a ‘wanting’ notion, except in a few residual contexts, limited to second person experiencers. These are exemplified in (16):

- (16) a. *Siéntate, Michael ¿qué gustas? Yo invito ¿Sólo café?*
 sit.down Mike what like.2SG I invite only coffee
¿Estás seguro?
 are.2SG sure
 ‘Sit down, Mike, what do you like/want? I’ll get this. Just a coffee? Are you sure?’
 (1979, Morales)

16. When dative *gustar* is meant to convey a notion of wanting or wishing, the *irrealis* dimension of the utterance shows up in the verbal morphology (subjunctive or conditional mood); cf. *me gustaría estar contigo* ‘I would like to be with you’ (CREA, 2002).

- b. *Puedes empezar cuando gustes.*
 may.2SG begin when like.2SG
 'You may start whenever you like/wish' (1980, Rojas)

By contrast, *gustar* with dative experiencer is used very frequently in present-day Spanish to refer to people's likes and dislikes (*no gustar*) in the mental world. It is important to note that the advancement of the new construction went hand in hand with a growing tendency for the dative experiencer to occupy the initial subject position, until the DAT/EXP – V – NOM/STIM word ordering pattern became more or less fixed. The word order preference makes sense from the point of view of the inherently greater prominence of the human experiencer relative to the stimulus, which, in the case of *gustar*, tends to be non-human. The Spanish language is particularly sensitive to the familiar Animacy Hierarchy, and most of its two-place construals involving a dative marked individual and an element of inanimate reference display the dative-first ordering (Melis & Flores, 2013). It is due to the presence of these sentences that Spanish partakes in the current discussion on the phenomenon of non-canonically marked, oblique, or quirky subjects.

5. The 'liking' verbs with a human stimulus

Before we conclude this paper, we will draw a very quick picture of the behavior of the Spanish 'liking' verbs in correlation with human stimuli. This is an area in which nowadays *amar*, *querer*, and dative *gustar* mostly converge, with closely interconnected meanings, yet they are separated by subtle differences. Consider these examples, which are fairly representative of their uses:

- (17) a. *Tè amo, Ave Eva; te amo como se ama*
 you.ACC love.1SG Ave Eva you.ACC love.1SG as REFL love
a Dios
 ACC God
 'I love you, Ave Eva; I love you as one loves God' (1988, Romero)
- b. *No sólo te quiero cielo mío, sino que te necesito*
 not only you.ACC love.1SG heaven mine but that you.ACC need.1SG
 'I don't just love you, my darling, I need you' (1994, Buero Vallejo)
- c. *No me gusta la gente que bebe*
 not me.DAT like.3SG the people.NOM REL drinks
 'I don't like people who drink' (1995, Sánchez-Ostiz)
- d. *Me gustas Tú sabes cuánto me gustas No lo*
 me.DAT like.2SG you know how.much me.DAT like.2SG not it
puedo evitar me muero de ganas de llevar-te a mi cama
 can.1SG avoid REFL die.1SG of desire of take-you to my bed

y pasar la noche contigo
and spend the night with.you

'I like you. You know how much I like you. I can't avoid it. I'm dying to take you to my bed and spend the night with you' (2002, Bayly)

Amar, as stated earlier, appears in contexts focusing on a deep and enduring love bond, associated with a notion of wanting to do good things for the person one loves. So we may say that *amar* represents the 'other-oriented' face of love (cf. Wierzbicka, 1992).

Querer comes to mean 'to love, be fond of, like' when construed with a human referent (never with a non-human stimulus; see note 10). This development is peculiar to Spanish within the Romance family, and has not been fully elucidated (Corominas & Pascual, 1981, s.v. *querer*). Although *querer* is often used as a near synonym of *amar*, (17b) suggests that it conserves overtones of desire, extending from its basic meaning. In this way, *querer*, much more clearly than *amar*, evokes the pleasure derived from being with the loved person and the desire for such contact; it emphasises the 'self-oriented' aspect of love (cf. Wierzbicka, 1992).

Finally, *gustar* with a human stimulus is employed in two types of environment. (17c) exemplifies the first type: the stimulus names a class or group of individuals, whether indefinite or generic, and the verb expresses the same kind of attitude one might have with respect to objects or situations. In the second type of context, the meaning of *gustar* moves closer to that of *amar* or *querer*, with the difference that *gustar* implicitly communicates that the experiencer finds the other person *sexually* attractive. In (17d) the usually implied meaning is made explicit. For an explanation of where the connotation of sexual desire comes from, we may appeal to the metaphor LUST IS HUNGER; THE OBJECT OF LUST IS FOOD discussed in Lakoff (1987; cf. Kövecses, 2000). The metaphor shows that people tend to conceptualise sexual desire in terms of the hunger drive and the eating process. It, therefore, provides us with a clue for *gustar*, whose relation to sexual desire takes us back, one more time, to the historically primary sense of the verb ('taste').

6. Conclusions

In this paper we focused on the development of a dative experiencer construction with Spanish *gustar* 'to please, like', which came to replace an older nominative experiencer use. Given that psych verbs within and across languages vary syntactically, the mentioned development may at first sight cause indifference. It becomes more interesting, however, when projected against the background of

the well-established diachronic tendency in European languages for oblique experiencer arguments to be replaced with nominative subjects, undoubtedly more 'regular' ones from a transitive syntax perspective.

The reverse movement analysed in this paper finds a partial explanation in two important factors. The first one points to the existence of a dative model, embodied in *placer* 'to please, like', to which *gustar* was attracted; the second factor relates to the preference within Spanish for the dative case marking pattern under study, contrary to many other European languages.

This said, the question as to why *gustar* evolved the way it did remains, since favorable circumstances cannot be equated with determining ones. Under a different scenario, *gustar* could have opposed resistance to the force of attraction and held on to its nominative experiencer, as other similar verbs did, like *amar* 'to love' and *querer* 'to want, like, love'. The central goal of this paper was to shed light on the semantic and pragmatic factors that played a role in motivating the extension of the pattern of *placer* to *gustar*.

With respect to these motivations, two instruments of analysis proved to be of invaluable help. First, we were able to verify the importance of probing into the source meaning of the verbal unit that is changing its formal behavior, in order to identify what semantic properties facilitated the coupling of the verb to the new syntactic template. And, secondly, we found evidence to support the hypothesis that grammatical changes arise within specific contexts of use, and require close attention to these contexts when one seeks to understand the conditions under which lexical units begin to transform (cf. emergence of preposition *pora/para* in Torres Cacoullós & Bauman, in this volume).

References

- Allen, C. L. (1995). *Case marking and reanalysis*. Oxford: Clarendon Press.
- Bardđal, J. (2009). The development of case in Germanic. In J. Bardđal & S. L. Chelliah (Eds.), *The role of semantic, pragmatic, and discourse factors in the development of case* (pp. 123–159). Amsterdam: John Benjamins. doi:10.1075/slcs.108.09bar
- Bauer, B. (2000). *Archaic syntax in Indo-European. The spread of transitivity in Latin and French*. Berlin: Mouton de Gruyter. doi:10.1515/9783110825992
- Bruce, K. L., & Bruce, L. P. (2010). Emotions in the Alambak lexicon. In K. A. McElhanon & G. Reesink (Eds.), *A Mosaic of languages and cultures: Studies celebrating the career of Karl J. Franklin* (pp. 38–59). Dallas, TX: SIL e-books.
- Bybee, J. L., & Pagliuca, W. (1987). The Evolution of Future Meaning. In A. Giacalone Ramat, O. Carruba & G. Bernini (Eds.), *Papers from the 7th International Conference on Historical Linguistics* (pp. 109–122). Amsterdam: John Benjamins. doi:10.1075/cilt.48.09byb
- Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar. tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

- Calderón Campos, M. (2018). Intersubjectification and textual emphasis in the use of *definite article + proper name* in Spanish. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 75–98). Amsterdam: John Benjamins. (in this volume) doi:10.1075/ihll.16.04cal
- Cole, P., Harbert, W., Hermon, G., & Sridhar, S. N. (1980). The acquisition of subjecthood. *Language*, 56, 719–743. doi:10.2307/413485
- Corominas, J., & Pascual, J. A. (1981). *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos.
- Croft, W. (1991). *Syntactic categories and grammatical relations. The cognitive organization of information*. Chicago, IL: University of Chicago Press.
- Dąbrowska, E. (1997). *Cognitive semantics and the Polish dative*. Berlin: Mouton de Gruyter. doi:10.1515/9783110814781
- Diewald, G. (2002). A model for relevant types of contexts in grammaticalization. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 103–120). Amsterdam: John Benjamins. doi:10.1075/tsl.49.09die
- Dixon, R. M. W. (1991). *A new approach to English grammar, on semantic principles*. Oxford: Clarendon Press.
- Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67, 547–619. doi:10.1353/lan.1991.0021
- Fischer, S. (2004). The diachronic relationship between quirky subjects and Stylistic Fronting. In P. Bhaskararao & K. Venkata Subbarao (Eds.), *Non-nominative subjects* (Vol. 1; pp. 192–212). Amsterdam: John Benjamins. doi:10.1075/tsl.60.11fis
- Harris, A. C., & Campbell, L. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620553
- Haspelmath, M. (2001). Non canonical marking of core arguments in European languages. In A. Y. Aikhenvald, R. M. W. Dixon & M. Onishi (Eds.), *Non-canonical marking of subjects and objects* (pp. 53–83). Amsterdam: John Benjamins. doi:10.1075/tsl.46.04has
- Heine, B. (2002). On the role of context in grammaticalization. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 83–101). Amsterdam: John Benjamins. doi:10.1075/tsl.49.08hei
- Hopper, P. J. (1991). On some principles of grammaticization. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization* (Vol. 1; pp. 17–35). Amsterdam: John Benjamins. doi:10.1075/tsl.19.1.04hop
- Ibarretxe-Antuñano, B. I. (1999). Polysemy and metaphor in perception verbs: A cross-linguistic study. (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh, United Kingdom.
- Kemmer, S. (1993). *The middle voice*. Amsterdam: John Benjamins. doi:10.1075/tsl.23
- Kövecses, Z. (2000). *Metaphor and emotion. language, culture, and body in human feeling*. Cambridge: Cambridge University Press.
- Kuno, S., & Kaburaki, E. (1977). Empathy and syntax. *Linguistic Inquiry*, 8, 627–672.
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: The University of Chicago Press. doi:10.7208/chicago/9780226471013.001.0001
- Langacker, R. W. (1991). *Foundations of Cognitive Grammar. Vol. 2. Descriptive application*. Stanford, CA: Stanford University Press.
- Lewis, Ch. T. (1889). *Latin dictionary for schools*. Oxford: Clarendon Press.
- Maldonado, R. (1999). *A media voz. Problemas conceptuales del clítico se*. Ciudad de México: Universidad Nacional Autónoma de México.

- Melis, Ch. (1998). Sobre la historia sintáctica de *gustar*. In C. García Turza, F. González Bachiller & J. J. Mangado Martínez (Eds.), *Actas del IV Congreso Internacional de Historia de la Lengua Española*, (Vol. 2; pp. 295–305). Logroño: Universidad de la Rioja.
- Melis, Ch. (1999). Los verbos *placer* y *pesar* en la Edad Media: la expresión ‘impersonal’ de las emociones. In F. Colombo Airoldi (Ed.), *El Centro de Lingüística Hispánica y la lengua española* (pp. 87–105). Ciudad de México: Universidad Nacional Autónoma de México.
- Melis, Ch. (2011). Del placer al gusto. Cambio semántico y continuidad sintáctica. Paper presented at the *II Coloquio Internacional Lenguas y Culturas Coloniales* (5–9 septiembre, 2011). Ciudad de México: Universidad Nacional Autónoma de México.
- Melis, Ch., & Flores, M. (2013). On the historical expansion of non-canonically marked ‘subjects’ in Spanish. In I. A. Seržant & L. I. Kulikov (Eds.), *Diachronic typology of non-canonical subjects* (pp. 163–184). Amsterdam: John Benjamins. doi:10.1075/slcs.140.08mel
- Moore, J., & Perlmutter, D. M. (2000). What does it take to be a dative subject? *Natural Language and Linguistic Theory*, 18, 373–416. doi:10.1023/A:1006451714195
- Næss, A. (2007). *Prototypical transitivity*. Amsterdam: John Benjamins. doi:10.1075/tsl.72
- Onishi, M. (2001). Non-canonically marked subjects and objects: Parameters and properties. In A. Y. Aikhenvald, R. M. W. Dixon & M. Onishi (Eds.), *Non-canonical marking of subjects and objects* (pp. 1–51). Amsterdam: John Benjamins. doi:10.1075/tsl.46.03oni
- Sadler, M. (2007). *Grammar in use across time and space. Deconstructing the Japanese ‘dative subject’ construction*. Amsterdam: John Benjamins. doi:10.1075/sidag.20
- Seefranz-Montag, A. von. (1984). Subjectless constructions and syntactic change. In J. Fisiak (Ed.), *Historical syntax* (pp. 521–553). Berlin: Mouton de Gruyter. doi:10.1515/9783110824032.521
- Sigurðsson, H. A. (2002). To be an oblique subject: Russian vs. Icelandic. *Natural Language and Linguistic Theory*, 20, 691–724. doi:10.1023/A:1020445016498
- Sweetser, E. E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620904
- Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. III; pp. 57–150). Cambridge: Cambridge University Press.
- Torres Cacoullou, R., & Bauman, J. (2018). Allative to purposive grammaticalization: A quantitative story of Spanish *para*. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 195–221). Amsterdam: John Benjamins. (in this volume) doi:10.1075/ihll.16.9tor
- Vázquez Rozas, V., & Rivas, E. (2007). Un análisis construccionista de la diacronía de *gustar*. In I. Ibarretxe-Antuñano, C. Inchaurrealde & J. Sánchez (Eds.), *Language, mind and the lexicon* (pp. 143–164). Frankfurt: Peter Lang.
- Wierzbicka, A. (1992). *Semantics, culture, and cognition. Universal human concepts in culture specific configurations*. Oxford: Oxford University Press.

Corpus

Real Academia Española. *Corpus diacrónico del español* (CORDE). Madrid: RAE. <<http://www.rae.es>>

Real Academia Española. *Corpus de referencia del español actual* (CREA). Madrid: RAE. <<http://www.rae.es>>

Postverbal subjects of unaccusative verbs in the history of Portuguese

Kristine Gunn Eide

ILOS/ISWOC, University of Oslo

In this paper, I trace a word order change from Old Portuguese (OP) to Modern European Portuguese (MEP) that affected subjects of unaccusative verbs and argue that these subjects lost their object-like position along two different paths: subjects that contain old information follow the same development as subjects of transitive verbs, while new information subjects develop independently. In OP they retained more of their object-like properties, while MEP subjects of unaccusatives pattern more closely with subjects of transitive verbs.

Keywords: unaccusative; information structure; word order; Portuguese; subject; postverbal; transitive; inergative; new information; old information

1. Introduction

The background for this article is the historical development of subject positions with unaccusative verbs in Portuguese.¹ In clauses with unaccusative verbs, the two possible word orders subject-verb (SV) and verb-subject (VS), as in (1) and (2), both meaning ‘João arrived’, are found in all periods of Portuguese.

- (1) *O João chegou*
the João arrived.3sg
‘João arrived’

1. I am very grateful to Ioanna Sitaridou and two anonymous reviewers for valuable comments on this paper.

- (2) *Chegou o João*
 arrived.3SG the João
 'João arrived'

The frequency of inversion, however, changes according to syntactic and information-structural criteria. While in Old Portuguese (OP; up to and including the 15th century) and Modern European Portuguese (MEP; 19th century to present) there is a statistical tendency towards the SV order, Classical Portuguese (CIP; 16th-18th c.) has a clear preference for VS.

The aim of this article is to disentangle syntax and information structure and to show how these interact during different periods in the history of the language. In Sections 2, 3 and 4, I describe theoretical assumptions and the methodology as well as current theories within this area of Portuguese historical syntax. In Section 5, I compare the position of subjects of unaccusative verbs with that of subjects and objects of transitive verbs taking into account their information-structural function. In Section 6, I provide statistical evidence for the changes that have taken place in order to illustrate the process.

2. Theoretical assumptions

I assume that subjects of unaccusative verbs, like objects of transitive verbs, are internal arguments to the verb (cf. Burzio, 1986): (1) transitive verbs: DP [V DP], (2) inergative verbs: DP [V] and (3) unaccusative verbs: [V DP]. In MEP the word order with unaccusative verbs depends on syntactic restrictions such as case and agreement (Ambar, 1992; Costa, 2004), information-structural restrictions, such as given and new information (Costa, 2004), and prosodic restrictions, such as sentence final (focus) stress (Costa, 2004; Frota, 2000). While MEP is basically a SVO language, it is heavily restricted by information-structural criteria. Costa (2004) shows how, in information-structurally marked contexts, MEP will mark old and new information through word order in the following way: new information is sentence final and has sentence focus stress. At the beginning of the sentence (before the verb) we will find old information, topics and, in unmarked contexts, new information, as illustrated by the question/answer pairs in (3)–(7):

- (3) A: *O que é que aconteceu?*
 the what is which happened.3SG
 'What happened?'
 B: *O Paulo partiu a janela*
 the Paulo broke.3SG the window
 'Paulo broke the window'

- B': #*Partiu o Paulo a janela*²
broke.3SG the Paulo the window
- B'': #*A janela, o Paulo partiu-a*
the window the Paulo broke.3SG-it
- (4) A: *O que fez o Paulo?*
what did.3SG the Paulo
'What did Paulo do?'
- B: *O Paulo partiu a janela*
the Paulo broke.3SG the window
'Paulo broke the window'
- (5) A: *O que partiu o Paulo?*
what broke.3SG the Paulo
'What did Paulo break?'
- B: *O Paulo partiu a janela*
the Paulo broke.3SG the window
'Paulo broke the window'
- (6) A: *Ninguém partiu nada*
no one broke.3SG nothing
'No one broke anything'
- B: *Partiu o Paulo a janela*
broke.3SG the Paulo the window
'Paulo broke the window'
- (7) A: *Quem partiu a janela?*
who broke.3SG the window
'Who broke the window?'
- B: *Partiu a janela o Paulo*
broke.3SG the window the Paulo
'Paulo broke the window'
- (8) A: *Quem é que partiu as janelas?*
who is which broke.3SG the windows
'Who broke the windows?'
- B: *Esta janela partiu o Paulo.*
this window broke.3SG the Paulo
'This window, Paulo broke'

2. The hashtag indicates that the answer, albeit grammatically correct, is infelicitous in the context.

(9) A: *A Ana viu o Paulo?*
 the Ana saw.3SG the Paulo
 'Did Ana see Paulo?'

B: *O Paulo, ela viu*
 the Paulo, she saw.3SG
 'Paulo, she saw'

(Examples from Costa, 2004, pp. 77–78)

In an unmarked context, as in (3), where the answer to the open question 'what happened?' is all new information, the only possible word order is SVO. The same word order is found in (4), where *o Paulo* is given information, and *partiu a janela* is new information, and in (5), where only the object *a janela* is new. In (6), both *o Paulo* and *a janela* are new information and we find the order VSO, whereas in (7), only *o Paulo* is new information and is sentence final, rendering the order VOS. As for preverbal objects, such as in (8) and (9), these have to have 'been referred to in the previous discourse and/or have some contrastive force.' (Costa, 2004, p. 80).

By comparison, an unmarked context similar to the one in (3), will yield both VS and SV with unaccusative verbs, as shown in (10).³ In other contexts, such as when *o João* is old information (or a topic), we find the SV order, as in (11). Conversely, when João is new information, we find VS, as in (12):

(10) A: *O que aconteceu?*
 what happened.3SG
 'What happened?'

B': *O João chegou.*
 the João arrived.3SG
 'João arrived'

B'': *Chegou o João*
 arrived.3SG the João
 'João arrived'

(11) A: *O que aconteceu com o João?*
 what happened.3SG with the João?
 'What happened to João?'

3. As Tortora (1996) has shown, verbs of inherently directed motion, such as *partir* 'to leave' and *chegar* 'to arrive' (*partire* and *arrivare*) can be divided into subgroups: one (the *arrivare* type) which has a covert implicit locative (goal) projected in its syntax, and another, (the *partire* type) which does not (source). As a result, in Italian, the unmarked word order for the *partire* type is SV, whereas for *arrivare* it is VS. When the *partire* type occurs with a postverbal subject, it has a contrastive interpretation. For Portuguese, on the other hand, a contrastive interpretation does not seem to be available at any stage in its history.

- B: *O João chegou*
 the João arrived.3SG
 'João arrived'
- (12) A: *Quem chegou?*
 who arrived.3SG
 'Who arrived?'
- B: *Chegou o João*
 arrived.3SG the João
 'João arrived'

When the subjects are postverbal, they can be analysed as remaining *in situ* as in (13) and when they are preverbal, they are in Spec-IP, as in (14), patterning with subjects of transitives and inergatives.

- (13) *Chegou o João* [IP V_i [VP_i S]]
 arrived.3SG the João
 'João arrived'
- (14) *O João chegou* [IP S_i V_j [VP_j i]]
 the João arrived.3SG
 'João arrived'

According to Costa (2004), objects in MEP may remain *in situ* or move out of VP, depending on their information status. Monosyllabic adverbs, such as *bem* 'well', provide a reliable test for marking the left edge of VP in Portuguese: arguments to the left of this adverb are situated above VP, while arguments to the right of this adverb remain within VP. In (15), where the language that Paulo speaks is new information, it remains within VP but moves out of VP when it is given information, as in (16).

- (15) A: *Há alguém aqui que fala bem francês ou inglês?*
 is anyone here that speaks.3SG well French or English
 'Is there anyone here who speaks French or English well?'
- B: *Não, mas o Paulo fala bem alemão*
 no but the Paulo speaks.3SG well German
 'No, but Paulo speaks German well'
- B': *#Não, mas o Paulo fala alemão bem*
 no, but the Paulo speaks.3SG German well
 '#No, but Paulo speaks German well'
- (16) A: *Como é que o Paulo fala francês?*
 how is that the Paulo speaks.3SG French
 'How does Paulo speak French?'

- B: *O Paulo fala francês bem*
 the Paulo speaks.3SG French well
 'Paulo speaks French well'
- B': *#O Paulo fala bem francês*
 the Paulo speaks.3SG well French
 '#Paulo speaks French well' (Examples from Costa, 2004, p. 41)

In MEP, sentence focus stress is sentence final. To avoid the subject of an unaccusative verb receiving focus stress, it must move out of VP, according to Costa, to Spec-IP, which is the canonical subject position in Portuguese (Ambar, 1992; Costa, 2004).⁴ However it is important to note that the preverbal subject, even when containing new information, is also in Spec-IP, as in example (10) in which word order is variable in unmarked contexts.

To sum up, the following factors determine word order in Portuguese sentences:

- a. Syntactic restrictions: the unmarked word order is SVO. Subjects of unaccusatives are ambiguous: they are object-like in the sense that they are internal arguments to the verb, they are subject-like because they move to get case.
- b. Information-structural restrictions: subjects can be sentence final only if they contain new information.
- c. Prosodic restrictions: these are closely linked to information structure in the sense that in a language with sentence final stress, old information is not likely to be sentence final.

These three types of restriction have been studied and to a large extent charted for MEP, as I have described here. With respect to OP and CIP, the prosodic restrictions are somewhat more difficult to describe, given the fact that no acoustic evidence is available. That should not prevent us from assuming that certain parts of a sentence had prosodic prominence. There is a universal tendency toward focus being associated with some sort of prosodic prominence, of which MEP is an example. I will assume that in both OP and CIP, as in MEP, sentence stress was sentence final, since it is in this position we systematically find new information.

4. Also, if it were a topic, it would move to a topic position in the left periphery, whether it contains new information or not.

3. Methodology

The data consists of 1150 declarative main clauses taken from texts from the 13th to the 20th century. The texts from the 16th to the 20th century have been analysed according to the respective authors' dates of birth and divided into periods of fifty years. This was not possible for the texts from the 13th to the 15th century as often neither the author nor the exact date of composition are known. In addition, statistical data from Fiéis (2003) are incorporated for OP.

The corpus consists of texts from the CIPM (*Corpus informatizado do Português Medieval*) (OP), the *Tycho Brahe Parsed Corpus of Historical Portuguese* (CIP), as well as a selection of 19th and 20th-century literary texts, and four editions of the Portuguese online newspaper *O Público Digital*. Subjects of unaccusative verbs have been classified according to their information value in the following way:⁵

- a. The subject is mentioned in the preceding text.
- b. The subject refers to something mentioned in the preceding text, as in example (17) where *o rebentamento* 'the blast' refers to the previously mentioned explosion.
- c. Part of the subject is mentioned in the preceding text, as in (18) in which the subject of *ocorrerá* 'will take place', i.e. *a abertura* 'the opening', refers to the previously mentioned *visitas* 'visits'. However, the other part of the subject, *ao público* 'to the public' is not the same as *privadas* 'private', rather it contrasts with it. The subject as a whole will thus be interpreted as contrastive focus at the same time as it introduces a new referent.
- d. The subject is not previously mentioned and does not refer to or contrast with anything in the preceding text.

5. The following verbs have been used for this study, classified according to Levin and Rapaport Hovav's (1995) classifications. Verbs of inherently directed motion: *chegar* 'come', *entrar* 'enter', *sair* 'exit', *partir* 'leave', *tornar* 'turn, return', *voltar* 'return', *subir* 'rise', *cair* 'fall', *passar* 'pass', *fugir* 'flee'; verbs of existence: *existir* 'exist', *sobrar* 'remain', also non-existence *faltar* 'lack, be missing'; verbs of appearance: *surgir* 'surge', *começar* 'begin', *rebentar* 'burst', *levantar-se* 'rise', *originar-se* 'originate', *amanhecer* 'dawn', *nascer* 'be born'; verbs of occurrence: *acontecer/ocorrer/passar-se* 'happen', *seguir-se* 'follow'; verbs of disappearance: *desaparecer/sumir* 'disappear', *morrer/falecer* 'die', *acabar* 'end', *parar* 'stop'; externally caused verbs of change of state: break verbs: *britar/quebrar/quebrantar/romper* 'break'; other alternating change of state verbs: *queimar/arder* 'burn', *melhorar* 'improve', *abrir* 'open', *fechar, serrar* 'close', *encerrar* 'lock', *aquecer* 'become warm', *esfriar* 'become cold', *amarelecer* 'redden', *empalidecer* 'turn pale', *danar* 'hurt', *sanar* 'heal'.

- (17) *Seis pessoas morreram ontem na sequência da explosão de uma bomba num autocarro no Estado do Penjab, no norte da Índia. O reventamento, que feriu 25 outras pessoas, ocorreu quando a viatura circulava entre as cidades de Pathankot e Jammu, a capital da Caxemira indiana*
 ‘Six people died yesterday following a bomb explosion on a bus in the State of Punjab in the north of India. The blast, which injured another 25 people, **happened** when the vehicle was travelling between the cities of Pathankot and Jammu, the capital of Indian Kashmir’ (O Público Digital)
- (18) *Entretanto, prosseguem as obras de construção do edifício, desenhado pelo arquitecto Frank O. Gehry, situado na margem da ria da cidade basca. “O interior está terminado”, disse o porta-voz e assegurou que o resto ficará pronto nos primeiros dias de Setembro. O início das visitas privadas está previsto para o próximo dia 3 de Outubro, enquanto a abertura ao público ocorrerá duas semanas depois, no dia 18*
 ‘Meanwhile, the construction works on the building continue. The building was designed by Frank O. Gehry and is situated on the riverbank in the Basque city. “The interior is completed”, said the spokesperson and he confirmed that the rest will be ready by the first days of September. The first private visits are scheduled for October 3rd, while the opening to the public will take place two weeks later, on the 18th’ (O Público Digital)

The subjects of the first two subgroups, a and b, exhibit the same behaviour and have been classified as referring to old information, those of the latter groups, c and d, on the contrary, as referring to new information. This classification does not make use of the concepts topic and focus since they (in particular topic) are somewhat unclear categories, not only with regard to their numerous definitions within the linguistic literature, but also with regard to their realisation. It is difficult to separate topic from other types of old/non-focus information (see Frota, 2000, regarding the ‘fuzziness’ in topic pronunciation in MEP). The dichotomy given/new, although not always easy to disentangle in older texts, is a much more objective criterion since no word can be classified as old without having an antecedent in the text itself. The present paper will, however, not take into consideration contrast. Contrasted elements may appear in both preverbal and postverbal position as shown in (19), where they are underlined. Note that it is not clear exactly which factors are responsible for their distribution (but see Eide & Sitaridou, 2014 for a discussion as to how contrast can be disentangled from other types of information structure in older texts).

- (19) *E avéença averam tam grande antre sy que o
 and jealousy have.FUT.3PL so big between themselves that the
que hũu quiser eso quereram os outros e o
 which one want.FUT.SUBJ.3SG that want.FUT.3PL the others and the*

que os outros quiserem esso quererá elle
 which the others want.FUT.SUBJ.3SG that want.FUT.3SG he
 ‘And the jealousy between them will be so great that that which one wants
 will be desired by the others, and that which the others want will be desired
 by him’
(*Visão do Túndalo*)

The data from the Information Structure and Word Order Change in Romance and Germanic Languages Corpus (ISWOC), which has a more refined annotation than the one described above (see Bech & Eide, 2012,; PROIEL, 2011 for details), has been presented here in a simplified version in order to make the annotations comparable to the method for the CIP and the MEP texts.⁶

4. Previous descriptions of Portuguese historical syntax

Over the past two decades, many of the descriptions of Portuguese syntax have centered around the discussion of whether or not OP and CIP were V2 languages. In particular it is the large number of subject-verb inversions in these that has led some authors (Galves & Galves, 1995; Ribeiro, 1995; Salvi, 2004) to analyse them as such. The V1 and V3 sentences, as in (20) and (21) respectively, have led other authors to reject the V2 hypothesis and analyse Portuguese at all stages as an SVO language with information structural constraints (Fiéis, 2003; Martins, 1994; Rinke, 2007). Fiéis (2003) finds that the postverbal subjects in her corpus were mostly subjects of unaccusatives and concluded that the large number of inversions with a sentence-initial adverb, as in (22), were mostly due to unaccusativity. Inversion with transitive verbs was ascribed to information structural effects. Rinke (2007) also argues that in OP there is no evidence that the verb obligatorily moves to C, and that the word order variation can be accounted for by information structure, where new information subjects and objects remain *in situ* in the same way that has been described for MEP.

- (20) *E tornavã as serpentes aas almas e comi[a]n-has*
 and turned the serpents to-the souls and ate-them
 ‘And the serpents turned to the souls and ate them’
(CIPM, *Visão de Túndalo*)

6. ISWOC is a research project of the University of Oslo, Norway. The OP part of this corpus consists of texts from CIPM which have been annotated for morphology, syntax and information structure.

- (21) *Entõ a alma começou de chorar*
 then the soul started.3SG to cry
 ‘Then the soul started to cry’ (CIPM, *Visão de Túndalo*)
- (22) *E entõ apareceo o angeo*
 and then appeared.3SG the angel
 ‘And then the angel appeared’ (CIPM, *Visão de Túndalo*)

As mentioned, CIP has also been analysed as a V2 language (Galves & Galves, 1995; Galves & Sousa, 2005) as there is a sharp rise in postverbal subjects in the 16th century. Moreover, the frequency of all types of verbs in clauses with an XVS structure, such as (23)–(28), regardless of whether the subject contains old or new information, supports the V2 hypothesis. In the following examples, information status is indicated in parenthesis.

Inergative:

- (23) *E com êle mandou J. G. Sarmento*
 and with him commanded.3SG J.G. Sarmento(NEW)
 ‘And J. G. Sarmento held the command with him’
 (Diogo do Couto, *Décadas*)

Transitive:

- (24) *Isto nos afirmou muito um homem Polaco, chamado*
 this to-us affirmed.3SG much a man Polish(NEW) called
Gabriel [...]
 Gabriel [...]
 ‘And a Polish man called Gabriel strongly affirmed this [...]’
 (Diogo do Couto, *Décadas*)
- (25) *Tanto que foi de dia, mandou Dom Estevão recado*
 as soon as was day, sent.3SG Dom Estevão(OLD) message
aos Vereadores [...]
 to-the town-councillors [...]
 ‘At the break of day, Dom Estevão sent a message to the town
 councillors [...]’ (Diogo do Couto, *Décadas*)

Reflexive:

- (26) *Com isto se despedio Belchior Fernandes*
 with this REFL left.3SG Belchior Fernandes(OLD)
 ‘With this Belchior Fernandes took his leave’ (Diogo do Couto, *Décadas*)

Predicative:

- (27) *E assim era êste bárbaro tão afeiçoado aos*
 and thus was this barbarian(OLD) so friendly towards-the

Cristãos, que [...]

Christians, that [...]

‘And so this barbarian was so friendly towards the Christians, that [...]’

(Diogo do Couto, *Décadas*)

Unaccusative:

- (28) *Aqui surgiu a Armada, e se deteve*
 here appeared.3sg the Armada(OLD) and REFL stayed
trinta e dois dias
 thirty-two days

‘The Armada appeared here and stayed for thirty-two days’

(Diogo do Couto, *Décadas*)

We do, however, find V3 clauses (29) as well as V1 (30).

- (29) *Além disto, eu tenho para mim que [...]*
 aside from-this, I(OLD) hold.1sg for me that [...]
 ‘Besides, I am of the opinion that [...]’

(Francisco Rodrigues Lobo, *Côrte Na Aldeia e Noites de Inverno*)

- (30) *Começou el-Rei a igreja de Vicente*
 began.3sg the-King(OLD) the Church of Vincent
 ‘The King started building St.Vincent’s Church’

(Frei Luís de Sousa, *Frei Bartolomeu dos Mártires*)

As for MEP, a sentence-initial adverb will not influence the subject position, and the unmarked word order with transitive verbs will be as in (31) and with unaccusatives as in (32).

- (31) *Ontem o João comprou o livro*
 yesterday the João bought.3sg the book
 ‘Yesterday João bought the book’

- (32) *Ontem o João chegou. Ontem chegou o João*
 yesterday the João arrived.3sg yesterday arrived.3sg the João
 ‘Yesterday John arrived’

For the purposes of the present paper, which focuses on the postverbal part of the sentence, the position of the verb, whether it moves to C° or, as Martins (1994) suggests, to the highest projection in the IP domain (ΣP), will not be discussed further. It is clear, however, that both topicalisation and contrastive focus fronting occurred in OP (Martins, 2011) and that subjects of unaccusatives were also affected by these constraints. Topics are associated with old information and are expected to occur in the left periphery. Contrastive focus fronting seems to have been less restricted in OP than in MEP, as mentioned by Costa and Martins (2011) who point to several semantic restrictions on arguments that undergo this type of fronting.

5. Subject and object positions in the history of Portuguese

Given the basic assumption that unaccusative subjects have both subject and object properties, I will present an overview of the subject and object positions in the diachrony of Portuguese, in combination with information structure (cf. Schulte, in this volume for the diachrony of pronominal subjects in Spanish).

5.1 Subject positions with transitive and inergative verbs

In Table 1, we see that subjects, whether old or new, can occur both pre- and post-verbally in OP (33)–(36) and CIP (37)–(40).⁷ The main historical development is the loss of a postverbal position for subjects that are old information in MEP. This change happens in the transition from CIP to MEP.

Table 1. Subject positions with transitive and inergative verbs

		OP	CIP	MEP
VS	Subject=Old	√(33)	√(37)	×
	Subject=New	√(34)	√(38)	√
SV	Subject=Old	√(35)	√(39)	√
	Subject=New	√(36)	√?(40)	√

OP

- (33) *E tendeo entō o angeo a mão e predeo*
 and reached-out.3SG then the angel(OLD) the hand and took
hũu daquelles dyabóos
 one of-those devils

'And the angel then reached out his hand and grabbed one of those devils'

(*Visão de Túndalo*)

- (34) *Tal pēna merecem os matadores e*
 such punishment deserve.3PL the killers and
os cōsentidores
 the accomplices(NEW)

'Such punishment is what murderers and their accomplices deserve'

(*Visão de Túndalo*)

7. In examples (37) and (39), the verbs *começar* 'begin' and *entrar* 'enter' are used as transitive verbs and take a direct object, *a igreja* 'the church' and *os vallos* 'the valleys' respectively.

- (35) *E desi a cabo de pouco aquella besta deitou-a*
 and from-there at end of little that beast(OLD) threw.3SG-her
de ssy ē fundo do lago
 from herself to bottom of-the lake
 ‘And then, after a little while, the beast threw her to the bottom of the lake’
 (*Visão de Túndalo*)
- (36) *E senhor entoncos todos reix ficarom os giolhos*
 and Lord then all-the kings(NEW) put the knees
ante ty
 before you
 ‘And Lord, then all kings kneeled before you’ (*Morte de S. Jerónimo*)

CLP

- (37) *Começou el-Rei a igreja de Vicente*
 began the-King(OLD) the church of Vincent
 ‘The King started building St. Vincent’s church’
 (Frei Luís de Sousa, *Frei Bartolomeu dos Mártires*)
- (38) *Isto nos afirmou muito um homem Polaco, chamado*
 this to-us affirmed.3SG much a man Polish(NEW) called
Gabriel [...]
 Gabriel [...]
 ‘And a Polish man called Gabriel strongly affirmed this [...]’
 (Diogo do Couto, *Décadas*)
- (39) *Os Turcos entrarom os vallos por duas partes*
 the Turks(OLD) entered.3PL the ditches from two sides [...]
 ‘The Turks entered the ditches from two sides’ (Diogo do Couto, *Décadas*)
- (40) *Faleceo êste Rei sem deixar filhos, e os povos*
 died this king without leave.INF children and the people(NEW)
levantaram por Rei Ceidafim
 raised.3PL as king Ceidafim
 ‘This king died childless, and the people made Ceidafim king’
 (Diogo do Couto, *Décadas*)

5.2 Postverbal object positions with transitive verbs

Following Costa (2004) and for OP Martins (2011), I assume that certain adverbs, such as *bem* ‘well’ and *sempre* ‘always’, can be used as a diagnostic for VP boundaries in MEP as well as OP. The examples available with this type of indicator are few in the texts and it has not been possible to decide how information structure is relevant for object movement in OP. Martins (2011) proposes that the same rule

that applies for MEP objects of transitive verbs also works in OP: the object moves out of VP in order to avoid the sentence final focus stress. Example (41) and (43) show that the dichotomy of given/new information does not capture this rule in an absolute way.⁸ The object in these examples, in both cases a certain vineyard, has been mentioned in the preceding text. The motivation for the word order in (43) is not necessarily a defocusing of the object, it can also be seen as a way of focusing the adverb.

Table 2. Transitive verbs, postverbal object positions [Vi XP] [VP ti tj]/[Vi [VP ti XP]

		OP	CIP	MEP
V _[VP] O	Object=Old	√ (41)	?	×
	Object=New	√(42)	√(44)	√ (46)
V O _[VP]	Object=Old	√ (43)	√ (45)	√ (47)
	Object=New	?	?	×

The question marks in Table 2 indicate that, while no examples were found in these categories, the size of the material is not sufficient to draw the conclusion that they do not exist. If Martins (2011) is correct in assuming that objects move out of VP in order to avoid focus stress, we might fall upon examples with a similar structure as the one in (46) where the need to stress the adverb overrides the given/new dichotomy. In the present corpus, all examples such as (46) have previously mentioned subjects. However, this is also true for the examples with structures as in (41), where an old information object is preceded by the adverb *bem*. In fact this structure is just as common with old information subjects as the construction in (43). Nevertheless the examples (41) and (43) show that objects do occur *in situ* as well as outside of VP, even though the information structural status of the object in (41) is problematic.

OP

- (41) [...] *que adubedes bem a dita vjnha*
 [...] that manure.CONJ.2PL well the said vineyard(OLD)
 '[...] that you shall manure the said vineyard well'
 (Legal document 1394 CIPM DN150)

8. Rather, the given/new dichotomy should be seen as a transparent way of establishing the main IS structures such as topic and focus. There is a correlation topic-old information-subject and focus-new information-object. So while focus, new information and object will normally coincide, they do not do so with necessity.

- (42) [...] *os que ham senpre vōotade de mal fazer*
 [...] those who have.3PL always desire(NEW) to evil do
 ‘[...] those who always have a desire to do evil’ (Visão de Túndalo)
- (43) *uos [...] auedes a laurar esta vinha bẽ*
 you [...] have.2PL to work thi vineyard(OLD) well
 ‘you shall work this vineyard well’ (Legal document 1305, CIPM CHP108)

CIP

- (44) [...] *na qual vinha hun Bramene que falava muyto bem*
 [...] in which came.3SG a brahmin who spoke.3SG very well
Portuguez
 Portuguese(NEW)
 ‘In which [boat] came a brahmin who spoke Portuguese very well’
 (Fernão Mendes Pinto, *Peregrinação*)
- (45) *Dom Jorge recebeo o Mathias bem*
 Dom Jorge received.3SG the Mathias(OLD) well
 ‘D. Jorge received Mathias well’ (Diogo do Couto, *Décadas*)

MEP

- (46) *Paulo fala bem francês*
 Paulo speaks well French(NEW)
 ‘Paulo speaks FRENCH well’ (Costa, 2004, pp. 40–41)
- (47) *Paulo fala francês bem*
 Paulo speaks French(OLD) well
 ‘Paulo speaks French WELL’ (Costa, 2004, pp. 40–41)

The examples that were found in CIP, such as (44) and (45), indicate a defocusing phenomenon here as well, similar to the one we have seen in MEP in (15) and (16), repeated here as (46) and (47). There seems to be a relatively stable pattern in CIP and MEP where objects move above VP delimiters such as *bem* when they contain old information and are not expected to carry focus stress, and remaining below them when they provide new information and are focussed.

5.3 Subject order with unaccusative verbs

In the case of unaccusative verbs, we observe two changes from OP to MEP in Table 3: VS order with old information subjects disappears, while SV order with new information subjects appears. In other words, unlike subjects of transitive verbs, subjects of unaccusatives do not appear preverbally if they are new information in OP.

Table 3. Subject-Verb orders with unaccusative verbs in declarative main clauses

		OP	CIP	MEP
VS	Subject=Old	√(48)	√(51)	×
	Subject=New	√(49)	√(52)	√
SV	Subject=Old	√(50)	√(53)	√
	Subject=New	×	√(54)	√

OP

- (48) *E logo chegarō os dyaboos*
and at-once came.3PL the devils(OLD)
'And the devils came at once' (CIPM, *Visão de Túndalo*)
- (49) *E pella boca della sayam muy grandes*
and through-the mouth of-hers exited.3PL very big
chamas de fogo
flames(NEW) of fire
'And huge flames shot out of her mouth' (CIPM, *Visão de Túndalo*)
- (50) *Então o angeo desapareceo*
then the angel(OLD) disappeared.3sg
'Then the angel disappeared' (CIPM, *Visão de Túndalo*)

CIP

- (51) *E de então ficou Mestre Micael Angniolo em*
and from then remained.3sg Master Michael Angniolo(OLD) in
Roma até agora
Rome until now
'And from then and until now, Master Michael Agniolo remained in Rome'
(Fr. Holanda, *Da pintura antiga*)
- (52) *Não lhe faltava diligência e cuidado nacido*
not to-him lacked.3sg diligence and carefulness(NEW) born
da virtude
of-the virtue
'He did not want for virtuous diligence and care'
(Frei Luís de Sousa, *Bertolameu*, 1 II, p. 17)
- (53) *O Governador chegou à praia com bem*
the governor(OLD) arrived.3sg to-the beach with well
trabalho [...]
work [...]
'The governor arrived at the beach with much difficulty [...]'
(Diogo do Couto, *Décadas*)

- (54) *Duas grandes dificuldades me ocorreram acerca*
 two big difficulties(NEW) to-me happened.3PL regarding
destes casamentos [...]
 of-these weddings [...]
 ‘Two big difficulties occurred to me regarding these weddings [...]’
 (António Brandão, *Da Monarchia Lusitana*)

Unaccusative verbs with an explicit subject and a VP delimiter are quite scarce.⁹ From example (55), we can draw the conclusion that subjects providing new information may remain within VP in OP. From a corpus of CIP of around 5000 main and subordinate clauses with unaccusative verbs, the majority of which did not have an expressed subject, it has only been possible to find one single example which is clearly outside VP, given in (56).

- (55) [...] *porque jazia ē ella sempre a neve*
 [...] because lay.3SG on her always the snow(NEW)
 ‘[...] because the snow always covered it [the mountain]’ (CGE 12, fol. 9a)
- (56) [...] *porque lhe não parou o cavalo bem*
 [...] because for-him not stopped.3SG the horse(NEW/OLD) well
 ‘[...] because his horse didn’t stop’ (Diogo do Couto, *Décadas*)

Given the ambiguous nature of the information value of the subject in the example in (56) (it can be seen as new, since the horse has not been mentioned, or grouped with old, since a horse can be inferred from the knight that rides it), this example only shows us that the subject has left the VP – it does not show us whether it is for information structural reasons or syntactic ones.

In this section, I have given an overview of the main changes in Portuguese, within the given/new dichotomy: old information subjects no longer occur post-verbally, whether they are transitive or unaccusative. New information subjects of unaccusatives are not found in OP in preverbal position. They are found in CIP, though in limited numbers, as will be seen in Section 6, Table 4, and are frequent in MEP. In the next section, I present the statistical material that outlines the development of the word orders.

9. Considering the semantic properties of the two groups of unaccusative verbs that provide the largest number of examples, verbs of inherently directed motion and verbs of existence and appearance, this is not surprising. These verbs are not likely to occur with an adverb of manner such as *bem*, which is more likely to describe an activity.

6. Statistical analysis

The graph in Figure 1 illustrates the statistical change in the word order with unaccusative verbs. In OP around one third of the subjects are postverbal, which is not very far from the frequency found in MEP. The period in the middle shows the opposite pattern: two thirds of the subjects are postverbal.

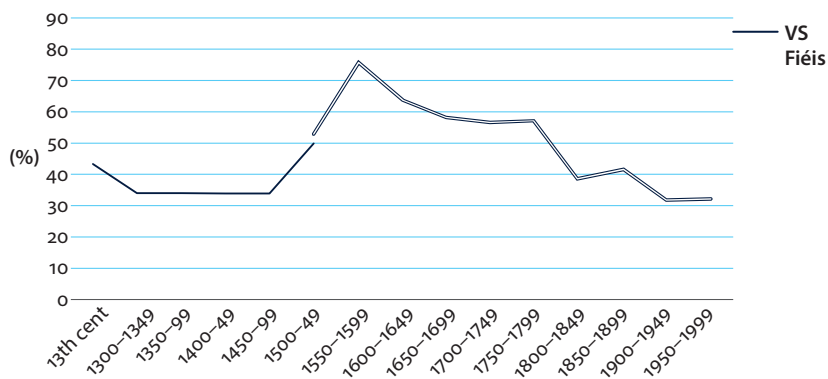


Figure 1. Percentage of postverbal subjects of unaccusative verbs in Portuguese (data are from Fiéis, 2003, and Eide, 2006)

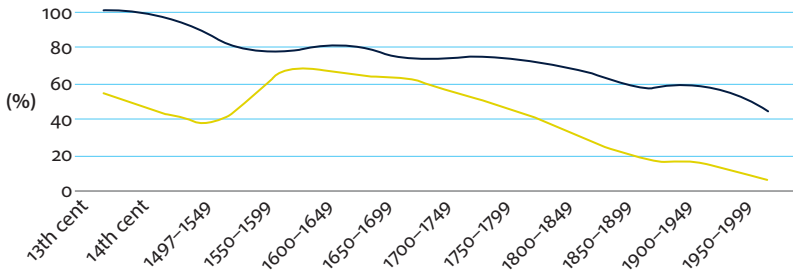
The same pattern as the one we see in Figure 1 emerges with transitive and inergative verbs: Schellert (1958) describes a sharp rise in postverbal subjects from the 13th/14th century to the 16th century, and Sousa (2004) observes a change around the end of the 17th/beginning of the 18th century where SV becomes more prominent. Galves and Sousa (2005) see this as a result of a change from preverbal topic positions to preverbal subject positions ($[\text{topic} - \text{verb} - \text{XP}] > [(\text{XP}) - \text{subject} - \text{verb}]$). They appear to assume that a reanalysis has taken place as follows, where the topic position in Spec-CP has been reanalysed as a canonical subject position: $[_{\text{CP}} \text{XP V} [\dots]] > [_{\text{IP}} \text{S V} [\dots]]$. Transitive verbs appear with more postverbal subjects in ClP (Eide, 2006) than in either of the two other periods, even though these subjects are at all times less likely to be postverbal than the subjects of unaccusative verbs. Table 4 shows the results of an analysis of VS and SV orders in combination with old and new information. No examples of preverbal new information subjects were found in the texts from 13th and 14th centuries. In the MEP corpus, however, there is a slight preference for the SV order even with new information subjects.

Figure 2, which is based on the numeric values in Table 1, illustrates the way only old information subjects are affected by the change in the 16th century. The figure shows a decrease in VS with both old and new information, but the decrease affects the old information to a much greater extent than the new information. It

Table 4. Total number of pre- and postverbal subjects of unaccusative verbs with old and new information

	13th c.	14th c.	1497– 1549	1550– 1599	1600– 1649	1650– 1699	1700– 1749	1750– 1799	1800– 1849	1850– 1899	1900– 1949	1950– 1999
VS Old	4	32	35	31	49	16	19	25	13	13	5	10
New	4	53	32	62	56	20	30	36	22	43	15	86
SV Old	4	31	51	16	27	10	17	24	33	66	32	153
New	0	0	7	12	13	7	11	14	11	32	11	108

is the subjects containing old information that pattern with those of the transitive verbs. In fact, it is these subjects only that account for the sharp increase in preverbal subjects in Figure 1. The new information subjects follow an independent pattern of becoming gradually more preverbal.

**Figure 2.** Information value of postverbal subjects of unaccusative verbs in main clauses

In Figure 2, the blue line shows the percentage of the new information subjects that are VS, and the yellow line, on the other hand, indicates the percentage of the old information subjects that are VS. The statistics are based on Eide (2006; 16th–20th centuries) and from the ISWOC-CIPM corpus.¹⁰ The examples from the 13th century are too few to be of any statistical significance. However, they do point to the same pattern that we find in the 14th century, where subjects that contain new information do not appear in preverbal position.

10. The ISWOC corpus, represented in Table 4 and Figure 2, so far suggests an even stronger preference for VS in OP than Fiéis' data do for the same period in Figure 1. The numbers and figures for OP in Table 4 and Figure 2 do not correspond directly to the statistics in Figure 1. Fiéis' data are from the same texts but from a larger part of the CIPM corpus and are thus more reliable from a statistical point of view. The discrepancies with regard to subject positions may be caused by the text selection.

7. Summary and conclusions

Based on the data above, three possible landing sites for the subjects of unaccusatives can be sketched:

- a. V [XP Subject *in situ*
- b. XP V [Subject in Spec-IP (or higher)
- c. V XP [Subject either scrambled (like an object) or in Spec-IP and verb moved to a higher projection

There has been a gradual drift towards SV order from OP/CIP to MEP. The change in the position of the subjects that contain old information corresponds to the changes that affect the subjects of transitive verbs, as described by Schellert (1958) and Sousa (2004). There was a sharp increase of VS in the 16th century followed by a decline around the turn of the 18th century. Given the general correlation between subjects, topics and old information, the fact that old information subjects of unaccusatives pattern with the subjects of the transitives is not surprising, since subjects of transitive verbs usually coincide with topics and are old information. Unaccusative verbs, on the other hand, are often used to introduce new referents and like objects of transitive verbs often correlate with focus and new information. In a discourse-driven language, we would expect subjects of unaccusatives to behave according to their information status. However, with the disappearance of a postverbal position for old information subjects, as in c, that took place between CIP and MEP, came a consolidation of Spec-IP as the canonical subject position it has today (Ambar, 1992; Costa, 2004). While Portuguese is still somewhat discourse-driven, subjects in unmarked transitive sentences, as well as topicalised subjects, are preverbal.

The gradual drift shown by the new information subjects then seems like a grammaticalisation process in which word order becomes more important in establishing grammatical relations, where the subject, even when new, appears preverbally. A reflex of this can be seen in the lack of agreement between subject and verb that is found in colloquial MEP in sentences like (57).

- (57) *Fechou muitas fábricas*
 closed.3SG many.PL factory.PL
 ‘Many factories closed’ (Costa, 2004, p. 116)

Such lack of agreement is not possible with preverbal subjects, as in (58), with postverbal subjects of transitives and inergatives, see (59), nor with postverbal subject pronouns, as in (60).

- (58) **Muitas fábricas fechou*
 many.PL factory.PL closed.3SG
- (59) **Cantou os meninos*
 sang.3SG the.PL boys.PL
- (60) **Chegou eles*
 came.3SG they.NOM.3PL

The lack of agreement in (57) shows that the single argument of the unaccusative verb has none of the canonical subject properties of Portuguese (agreement and preverbal position). In fact, it has the properties of an object: it lacks agreement and is in postverbal position. As expected, no examples of this pattern were found in OP and CIP. However, since this is only non-conclusive negative evidence, we cannot conclude that lack of agreement between the verb and postverbal subjects of unaccusatives is a direct result of the consolidation of the preverbal subject position. Nevertheless, it does point to the object properties associated with postverbal arguments.

The increase in preverbal subjects that provide new information has been gradual: from not being found in OP to around 50% in MEP. The fact that these subjects do not follow the pattern of the old information subjects indicates that we are dealing with a separate process, although the exact manner in which the change takes place is still not clear.

References

- Ambar, M. (1992). *Para uma sintaxe da inversão sujeito-verbo em Português*. Lisboa: Edições Colibri.
- Bech, K., & Eide, K. G. (2012). The annotation of morphology, syntax and information structure in a multilayered diachronic corpus. In F. Mambrini, M. Passarotti & C. Sporleder (Eds.), *Proceedings of the ACRH Workshop, Heidelberg, 5 Jan. 2012*. JLCL, 26(2), 13–24.
- Burzio, L. (1986). *Italian syntax. A Government-Binding approach*. Dordrecht: Reidel.
- Costa, J. (2004). *Subject positions and interfaces: The case of European Portuguese*. Berlin: Mouton de Gruyter.
- Costa, J., & Martins, A. M. (2011). On focus movement in European Portuguese. *Probus*, 23, 217–245. doi:10.1515/prbs.2011.006
- Eide, K. G. (2006). *Word order structures and unaccusative verbs in Classical and Modern Portuguese. The reorganisation of Information Structure*. (Unpublished doctoral dissertation). University of Oslo, Oslo, Norway.
- Eide, K. G., & Sitaridou, I. (2014). Contrastivity and information structure in Old Ibero-Romance languages. In K. Bech & K. G. Eide (Eds.), *Information Structure and syntactic change in Germanic and Romance languages* (pp. 377–412). Amsterdam: John Benjamins. doi:10.1075/la.213.14eid

- Fiéis, M. A. (2003). *Ordem de palavras, transitividade e inacusatividade. Reflexão teórica e análise do Português dos séculos XIII a XVI*. Lisboa: Universidade Nova de Lisboa.
- Frota, S. (2000). *Prosody and focus in European Portuguese: Phonological phrasing and intonation*. New York, NY: Garland.
- Galves, Ch., & Galves, A. (1995). A case study of prosody driven language change. From Classical to Modern European Portuguese. Retrieved from <http://www.tycho.iel.unicamp.br/~tycho/prfpml/fase1/papers/lang_change.pdf>
- Galves, Ch., & Paixao de Sousa, M. C. (2005). Clitic placement and the position of subjects in the history of Portuguese. In T. Geerts, I. van Ginneken & H. Jacobs (Eds.), *Romance languages and linguistic theory 2003. Selected papers from 'Going Romance' 2003, Nijmegen, 20–22 November* (pp. 93–107). Amsterdam: John Benjamins. doi:10.1075/cilt.270.06gal
- Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity at the syntax-lexical semantics interface*. Cambridge, MA: The MIT Press.
- Martins, A. M. (1994). *Clíticos na história do Português*. (Unpublished doctoral dissertation). Universidade de Lisboa, Lisbon, Portugal.
- Martins, A. M. (2011). Scrambling and information focus in Old and Contemporary Portuguese. *Catalan Journal of Linguistics*, 10, 133–158.
- PROIEL. (2011). Guidelines for the annotation of givenness. The PROIEL Project. Retrieved from <folk.uio.no/daghaug/info_guidelines.pdf>
- Ribeiro, I. (1995). Evidence for a Verb-Second Phase in Old Portuguese. In A. Battye & I. G. Roberts (Eds.), *Clause structure and language change* (pp. 110–139). Oxfordland: Oxford University Press.
- Rinke, E. (2007). *Syntaktische Variation aus synchronischer und diachronischer Perspektive*. Madrid & Frankfurt: Iberoamericana & Vervuert.
- Salvi, G. (2004). *La formazione della struttura di frase romanza: Ordine delle parole e clitics dal latino alle lingue romanze antiche*. Tübingen: Max Niemeyer.
- Schellert, D. (1958). *Syntax und Stilistik der Subjektstellung im Portugiesischen*. Bonn: Romanisches Seminar an der Universität Bonn.
- Schulte, K. (2018). On the position of overt subjects in infinitival clauses in Spanish and Portuguese: Pragmatic, semantic and frequency-based motivations. In M. Bouzouita, I. Sitari-dou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 173–194). Amsterdam: John Benjamins. (in this volume) doi:10.1075/ihll.16.08sch
- Sousa, M. C. P. de. (2004). *Língua barroca: Sintaxe e história do português dos seiscentos*. (Unpublished doctoral dissertation). State University of Campinas, Campinas, Brazil.
- Tortora, C. (1996). Two types of unaccusatives: Evidence from a Northern Italian dialect. In K. Zagana (Ed.), *Grammatical theory and Romance languages. Selected papers from the 25th Linguistic Symposium on Romance Languages (LSRL XXV), Seattle, 2–4 March 1995* (pp. 251–262). Amsterdam: John Benjamins. doi:10.1075/cilt.133.20tor

Corpus

Texts from the CIPM (*Corpus Informatizado do Português Medieval*):

13th century: Afonso X (parts), *Foro Real*; *Documentos Notariais* HGP102, 105, 107, 108, 139, 140, CHP 012, 013, 016, 020, 031, 033, 035, 038, 042, 043, 044, 047, 057, 058, 064, 067, CA 002, 031.

13th/14th century: *Vidas de Santos de um Manuscrito Alcobacense*.

14th century: *Crónica geral de Espanha* (parte 1, 1–13).

Texts from the *Tycho Brahe Parsed Corpus of Historical Portuguese*:

D. João III, *Cartas*; Diogo do Couto, *Décadas*, Quinta Decada, Livro oitavo, cap. 9, Livro décimo, cap. 11; Frei Luis de Sousa, *A Vida de D. Frei Bertolameu dos Mártires*, Livro I, cap. I–XIV; João de Barros, *Gramática da língua portuguesa* (from the beginning of the text until the use of tilde.); Fernão Mendes Pinto, *Perigração*, cap. I–XIV; Francisco de Holanda, *Da Pintura Antiga*, cap. I–XXIII; Manuel Pires de Almeida, *Poesia e pintura*; Manuel da Costa, *Arte de Furtar*; António Brandão, *Monarchia Lusitana*, Livro X, cap. I–IV; F. Rodrigues Lobo, *Corte na Aldeia*, Diálogo I + II; Manuel Severim de Faria, *Discursos vários políticos*, Carta ao leitor, Discurso primeiro; Manuel de Galhegos, *Gazeta*, Novembro-Dezembro 1941; António Vieira, cartas, II–IV, VIII–XI; António Vieira, Sermão, Primeira Domingo do Advento Prégado na Capella Real, 1652, I–IV; F. Manuel De Melo, *Cartas Familiares*, 6–16; F. Manuel de Melo, *Tácito Português*, Livro segundo; António das Chagas, *Cartas Espirituais*, 1–9; Marquesa de Alorna, *Cartas e outros Escritos*; José Daniel Rodrigues da Costa, *6 Entremeses de Cordel*; Correia Garção, *Dissertações*, 1–3, *Oração Quarta*; Almeida Garrett, *Viagens na minha terra*, cap. I–VII, cartas 1–41; Matias Aires, *Reflexão sobre a vaidade dos Homens*, cap. 1–50; António da Costa, cartas I–V; Cavaleiro de Oliveira, cartas 1–4; Diogo Ignácio de Pina Manique, *Pina Manique e a Universidade de Coimbra*; *Cartas do Intendente e de José Rodrigues Lisboa para o Doutor Francisco Montanha* (seleção e notas de Lígia Cruz), cartas 1–19; Ramalho Ortigão, *Cartas a Emília*, I–XIV; Marquês da Fronteira e d'Alorna, *Memórias do Marquês da Fronteira e d'Alorna*, Parte Primeira, 1802 a 1818, cap. I–II; Camilo Castelo Branco, *Amor de Perdição*, cap. I–V; Camilo Castelo Branco, *Maria Moisés* (Biblioteca Virtual de Autores Portugueses); J. M. Eça de Queiroz & J. P. Oliveira Martins, *Correspondência* (Texto introdutório de Paulo Franchetti. Fixação do texto, notas e comentários de Beatriz Berrini), Campinas, Editora da Unicamp, 1995, carta 1–21.

Additional literary texts from 19th and 20th centuries:

19th century: Abel Botelho, *O barão de Lavos*; Fialho de Almeida, *Contos*; Raúl Brandão, *Os pescadores*; Raúl Brandão, *Os pobres*; Júlio Dantas, *Os serenins de Queluz*; Almada Negreiros, *O Cágado*; Almada Negreiros, *Direcção Única*; Florbela Espanca, *A paixão de Manuel Garcia*; Mário de Sá-Carneiro, *A confissão de Lúcio*; Jaime Cortesão, *As memórias da grande Guerra*.

20th century: Manuel Alegre, *Rosas vermelhas*; António Arnaut, *Os dois barbeiros*; Maria Ondina Braga, *Carta de amor*; José Viale Moutinho, *Lucas depois do credo*; Fernando Namora, *A mulher afogada*; Urbano Tavares Rodrigues, *Dias coloridos*; Miguel Torga, *O Leproso*; Francisco José Viegas, *Troca de correspondência*; Lídia Jorge, *António*; Maria Judite de Carvalho, *O diário de saudade*; Maria Isabel Barreno, *A solução*.

Newspaper *O Público Digital*: *O Público* online, 28 de Junho de 1999, 13 de Agosto de 1997, 12 de Agosto 1997, & 7 de Abril de 1997.

On the position of overt subjects in infinitival clauses in Spanish and Portuguese

Pragmatic, semantic and frequency-based motivations

Kim Schulte

Universitat Jaume I

A synchronic survey of prepositional infinitive clauses reveals that the position of the subject depends heavily on the preposition heading the infinitival clause; however, equivalent prepositions in Spanish and Portuguese do not necessarily trigger a preference for the same word order. The differences can best be explained if the overt subject infinitive is analysed not as a single construction, but as a cover term for a series of related but independent constructions that are, diachronically, subject to different semantic, pragmatic and analogical pressures. Subsequent entrenchment of the most frequent patterns can eventually cause the subject to appear either in pre- or post-infinitival position by default. This process affects Spanish more than Portuguese, where the availability of an ‘inflected infinitive’ brings the syntax of infinitival clauses more into line with that of finite ones.

Keywords: Spanish; Portuguese; subject; infinitive; word order; frequency; preposition; position

1. Introduction

Most Romance languages allow overt subjects in infinitival clauses to some extent, which makes them cross-linguistically somewhat unusual. Whilst it is not particularly uncommon for infinitival clauses to contain an overt agent, this agent tends to appear in an oblique case, as in the English “exceptional case-marking” construction (*I want him to go away*) and in Hungarian, where the agent appears in the dative case and the infinitive can, furthermore, optionally be inflected to agree with the agent in terms of person and number, as shown in example (1), taken from Kenesei (1994, pp. 18–19).

- (1) *Fontos volt [Péter-nek úsz-ni(-a)]*
 important was Peter-DAT swim-INF-3SG
 'It was important for Peter to swim'

The fact that forms termed “infinitives” can have a varying proportion of noun-like and verb-like features (and a clear distinction between infinitives and deverbal nouns is often elusive) offers an explanation why the agents of infinitives often take the same case as the agents of “activity nouns”, as in the case of the “fourth infinitive” in several Finnic languages (Christen, 2001, p. 519).

The Romance languages, however, have developed a series of constructions in which the specified or overt agent of an infinitive appears in the nominative case, thereby qualifying as a fully-fledged subject of the non-finite subordinate clause. As Mensching (2000, p. 37) points out, “almost all Romance languages license specified subjects in infinitive clauses (or did so in the past) in a considerable number of configurations [...]”.

Diachronically, this phenomenon is even more clearly visible, as some of the less peripheral varieties such as Italo- and Gallo-Romance had a wider range of constructions in which infinitives could have their own overt subject in the past than they do at present: their use is highly restricted in the modern language (Mensching, 2000, pp. 18–23). In Spanish, Portuguese and Romanian, on the other hand, there has been no overall decline in the use of infinitives with overt subjects; within adverbial clauses, this type of construction has, indeed, experienced an increase in these languages. Whilst the continued presence of such constructions in Romanian (cf. Schulte, 2007, pp. 317–320) is noteworthy, given the overall more restricted use of the infinitive in Romanian compared with other Romance languages, it will not be discussed further in this paper because the position of the overt subject is obligatorily postverbal throughout the documented history of Romanian.

In the Ibero-Romance languages, in contrast, the overt subjects in infinitival adverbial clauses can appear either pre- or postverbally; the aim of this paper is to identify the factors that have contributed to the diachronic changes and the differing present-day distribution of these constructions in Spanish and Portuguese. These two languages can be understood to represent two different types regarding the position of overt subjects in adverbial clauses: (i) Catalan and several other Ibero-Romance languages/varieties largely resemble the Spanish pattern in this respect; whilst (ii) Galician and Portuguese (which both stem from a single Galician-Portuguese parent language) have developed along a different path, in which the availability of person/number agreement on the infinitive is likely to have been a decisive factor. The object of this paper is to analyse these differences between Spanish and Portuguese, both synchronically

and diachronically, with the aim of identifying the factors that have motivated such a distinct development in two sister languages that are, generally speaking, otherwise structurally very similar.

2. Overt subjects in Spanish and Portuguese infinitival clauses

2.1 Present-day distribution

Both Spanish and Portuguese allow infinitives to have their own overt subject, either a full NP or a subject pronoun, in a variety of clause types.

(2) Subject clauses

a. Spanish

sería interesante [cant-ar yo ahora la Traviata]
 would.be interesting [sing-INF PN.NOM.1SG now the Traviata]
 ‘It would be interesting for me to sing the Traviata now’

(cf. Yoon & Bonet-Farran, 1991, p. 353)

b. Portuguese

*Seria interessante [eu ir a China faz-er
 algumas pesquisas]*
 would.be interesting [PN.NOM.1SG go-INF to China do-INF
 some investigations]

‘It would be interesting for me to go to China to do some research’

(Kujawski, 2012)

(3) Object clauses

a. Spanish

*El adorador siempre lamenta [no ser él
 el adorado]*
 the worshipper always regrets [not be-INF PN.NOM.3SG
 the worshipped]

‘The worshipper always regrets not being the one (who is) worshipped’

(Docal, 2012)

b. Portuguese

*O João lamenta [eles ter-em gastado esse
 dinheiro para nada]*
 the John regrets [PN.NOM.M.3PL have-INF-3PL spent that
 money for nothing]

‘John regrets their having spent all that money for nothing’

(cf. Rouveret, 1980, p. 76)

(4) Adverbial clauses

a. Spanish

*Mientras yo iba en camino, [sin yo
While I went on way [without PN.NOM.1SG
sab-er-lo], mi esposa estaba preparando las maletas
know-INF-PN.ACC.3SG] my wife was preparing the suitcases
'While I was on my way, without my knowing, my wife was packing the
suitcases'*

b. Portuguese

*Ela morreu poucos minutos [depois.de eu cheg-ar]
she died few minutes [after PN.NOM.1SG arrive-INF]
'She died a few minutes after I arrived'*

(5) Attributive clauses

a. Spanish

*Han comprado una mesa [para com-er cuatro personas]
have bought a table [for eat-INF four persons]
'They've bought a table at which four people can eat'*

b. Portuguese

*Resolv-eu fazer uma casa [para eles viv-er-em]
resolved make a house [for PN.NOM.M.3PL live-INF-3PL]
'He decided to build a house for them to live in'*

Examples (2)–(5) show that all basic types of infinitival subordinate clauses can contain an overt subject, in both Spanish and Portuguese. However, the syntactic restrictions on the Overt Subject Infinitive (OSI hereafter) in Spanish are often overstated in the literature. Skydsgaard (1977, p. 17), for instance, observes that Spanish allows the infinitive to have its own syntactic subject only in certain, very specific situations. Hernanz Carbó (1982, p. 357), Fernández Lagunilla (1987, p. 128), Morales de Walters (1988, p. 90), Yoon and Bonet-Farran (1991, p. 357), Torrego (1998, p. 209) and Ledgeway (2000, p. 120) all claim that Spanish does not license infinitive clauses with overt subjects in object (complement) position, which is shown to be inaccurate by examples such as (3a); it is, nevertheless, true that object clauses are the most restrictive syntactic environment for OSIs, as the overt subject has to be pronominal and coreferential with the subject of the main clause. The issue of coreferentiality between the subject of the main clause and the infinitive is a further source of inaccuracy in the description of syntactic restrictions on the OSI; Fernández Lagunilla (1987, p. 135) states that “the presence of the lexical subject depends not only on the infinitive not being governed, but also on there being no coreference”, and Ledgeway (2000, p. 123) claims that “the personal infinitive [is] restricted to marking non-coreferentiality and

never surfaces in contexts of coreferentiality, where subject reference is already independently recoverable". This could hardly be further from the truth, as the vast majority of overt subjects in Spanish infinitival clauses are coreferential with the subject of the main clause (Schulte, 2007, p. 158); for a clear identification of non-coreferential subjects in subordinate clauses, finite subordination is a readily available alternative.

Having established that the syntactic contexts in which overt subjects occur in infinitival clauses are not fundamentally different in the two languages under consideration, it appears pertinent to ask why it is that they are considered so typical of Portuguese, but not of Spanish. The answer is that it is mainly a matter of usage frequency; a statistical analysis, based on the diachronic corpora of the respective languages used in Schulte (2007, pp. 371–381), reveals that whilst more than a third of Portuguese infinitives have an overt subject, this is only the case for approximately 1% of Spanish infinitives.¹

2.2 The Portuguese inflected infinitive

A crucial difference between Spanish and Portuguese regarding the OSI constructions is the obligatory morphological agreement of the infinitive in the presence of an overt subject in Portuguese. Whilst the person/number inflections are null morphemes for the first and third person singular, they are visible in the second person singular and the plural forms, as shown in (6).

- (6) 1SG *para eu* *faz-er-Ø*
 for PN.NOM.1SG do-INF-Ø
 ‘for me to do’
- 2SG *para tu* *faz-er-es*
 for PN.NOM.2SG do-INF-2SG
 ‘for you to do’
- 3SG *para ele* *faz-er-Ø*
 for PN.NOM.3SG do-INF-Ø
 ‘for him to do’
- 1PL *para nós* *faz-er-mos*
 for PN.NOM.1PL do-INF-1PL
 ‘for us to do’

1. Throughout this paper, observations regarding the syntax of OSI constructions in Spanish and Portuguese will be based on this corpus, which covers the entire period between the 12th and the 20th century; it includes a variety of text types for most periods, though due to availability restrictions, the majority are literary, administrative, historical and religious texts.

2PL	<i>para vós</i>	<i>faz-er-des</i>
	for PN.NOM.2PL	do-INF-2PL
	‘for you to do’ ²	
3PL	<i>para eles</i>	<i>faz-er-em</i>
	for PN.NOM.3PL	do-INF-3PL
	‘for them to do’	

It should be noted that the presence of an overt subject requires the infinitive to be inflected for person and number, but the use of inflectional morphology on the infinitive does not trigger obligatory presence of an overt subject, as the inflection alone is frequently sufficient to clarify the identity of the subject/agent of the infinitival clause, be it coreferential or not. In Section 3.2.2 it will be argued that the availability of inflected infinitives without specified subjects is partly accountable for the differences in the development of personal infinitives in Spanish and Portuguese.³

In clauses *with* overt subjects, however, the difference between the Spanish and the Portuguese OSI constructions is superficial. The obvious distinction between Spanish and Portuguese, namely the fact that an overt subject obligatorily implies person/number agreement between the subject and infinitive in Portuguese, whilst the absence of any such agreement is just as predictable in Spanish, is purely formal and should not necessarily be expected to have any particular effect on parameters such as the position of the overt subject (pre- or postverbal) or its referentiality (coreferential or non-coreferential).

3. The position of the overt subject in relation to the infinitive

3.1 A general overview

Whilst the clause types in which overt subjects can occur do not differ fundamentally between Spanish and Portuguese, as shown in Section 2.1, a crucial difference is the position of the subject in relation to the infinitive, as clearly identified by Mensching (2000).

2. The second person plural is rarely used in Modern Portuguese; it is commonly replaced by the third person plural with the pronoun *vocês*: *para vocês fazerem*.

3. The term ‘personal infinitive’ is not used consistently in the literature: whilst Ledgeway (2000, p. 115) sees it as opposed to the inflected infinitive, Posner (1996, p. 165) equates the two. Here, the term will be understood to include all infinitives with overtly specified subjects, whether this specification is achieved by means of inflection or by the overt presence of the subject.

3.1.1 Spanish

[...] earlier stages of Spanish allowed preverbal subjects in addition to postverbal position, although the latter already prevailed in Old Spanish and its frequency is even higher in Classical Spanish. Unlike Italian, there were no restrictions in the environments where either position occurs. [...] From the 18th century onwards, preverbal subjects no longer occur in almost any infinitival environments. The only constructions where preverbal subjects are still grammatical today are adverbial clauses (cf. Gili Gaya 1985: 189; Fernández Lagunilla 1987: 127, 132–33). These cases are characteristic for the spoken language. (Mensching, 2000, p. 25)

The claim that overt subjects have always been predominantly post-infinitival in Spanish is somewhat weakened by Keniston (1937, p. 550), who observes “a fairly strong tendency to place the subject before the infinitive” in 16th century Spanish. Ledgeway (2000, p. 292) points out that Keniston provides only examples of pronominal pre-infinitival subjects, which would suggest that nominal subjects were restricted to the postverbal position in 16th century Spanish. Table 2 (see Section 3.3.1) confirms that even in Modern Spanish, preverbal pronominal subjects are not rare or exceptional, at least in certain environments.

3.1.2 Portuguese

[The] construction with a nominative subject admits postverbal subjects. An extremely striking syntactic difference, especially in comparison to Spanish, is the high frequency of preverbal subjects, [...] a fact that already seems to have been the case in the oldest medieval texts. In most constructions, both positions are possible, the postverbal one being rather marked and used for focalizing and emphasizing purposes. (Mensching, 2000, p. 28)

This preference for preverbal position does not apply to the subjects of epistemic and declarative verbs, which generally permit only post-infinitival subjects (cf. Ambar, 1994; Mensching, 2000, p. 28; Raposo, 1987, p. 98), though some examples of preverbal subjects can be found in Brazilian Portuguese, as in *Dos crimes, não creio ele ser capaz* ‘Of these crimes, I don’t believe him to be capable’.⁴

In the remainder of this paper, it will be attempted to explain how this difference in preferred subject position between Spanish and Portuguese came

4. This example is taken from the lyrics of *Pilatos II*, performed by heavy metal band *Imperial Metal*.

about, what difference the choice between pre- and post-infinitival subjects makes in each of the two languages, and why, somewhat surprisingly, some types of Spanish adverbial clauses buck the trend and favour preverbal overt subjects.

3.2 The position of pronominal subjects of infinitives in Portuguese adverbial clauses

3.2.1 *Modern Portuguese*

Table 1 presents a breakdown of the position of overt pronominal subjects in different types of prepositional adverbial clauses; the data was extracted from the 180 million word *Cetem Público* journalistic corpus. For the purpose of this study, the category “preposition” will be defined as including any lexical element or sequence that can link an infinitival adverbial clause to a main clause; this includes compound prepositions such as concessive *apesar de* ‘despite’ as well as temporal *ao* ‘when’.

Table 1. Position of pronominal subjects in prepositional adverbial clauses

Preposition	Preverbal subj. pronouns	Postverbal subj. pronouns	Postverbal subj. pronouns with an overt intensifier
<i>ao</i> ‘when’	0% (0)	100% (7)	71% (5)
<i>por</i> ‘by’	89% (646)	11% (71)	52% (37)
<i>para</i> ‘in order to’	93.5% (1310)	6.5% (85)	66% (56)
<i>apesar de</i> ‘despite’	95.8% (213)	4.2% (9)	44% (44)
<i>sem</i> ‘without’	97.5% (120)	2.5% (3)	100% (3)
<i>até</i> ‘until’	97.6% (168)	2.4% (4)	75% (3)
<i>antes de</i> ‘before’	97.7% (343)	2.3% (8)	88% (7)
<i>depois de</i> ‘after’	>99% (351)	<1% (3)	100% (3)

Firstly, the data in Table 1 confirms that there is a strong tendency for Portuguese overt subjects to precede the infinitive, as observed by Mensching (2000, p. 28). The exception to this pattern is infinitival adverbial clauses introduced by *ao*, which will be discussed below.

It can further be seen that there is a clear link between postverbal subject position and the presence of an accompanying overt intensifier. According to König’s (2001, p. 75) definition, “*Intensifiers* evoke alternatives to the referent of their focus and structure the set of referents under consideration (referent of the focus + set of

alternatives) in terms of center and periphery”. Intensifiers thus have the discourse pragmatic function of marking *contrastive focus*.⁵

The intensifiers used with post-infinitival subjects in Portuguese can be subdivided into two types. On the one hand, there are lexical intensifiers, as in *ele mesmo* ‘he himself’, *elas próprias* ‘they themselves’, which account for around 60% of all intensified overt subjects in post-infinitival position. The remaining 40% are intensified by being incorporated into intensifying constructions such as *ele e só ele* ‘he and nobody else’, which accounts for approximately 10% of intensified overt subjects. A more complex construction that could be described as cleft infinitival clause, used in about 30% of cases, is [*ser* + subject pronoun + *a* + infinitive], as in example (7).

- (7) *tendo a posse de bola [para ser-em eles a*
 holding the possession of ball [for be.INF-3PL PN.NOM.3PL to
dit-ar-em o ritmo [...] de jogo]
 dictate-INF-3PL the rhythm [...] of game]
 ‘keeping possession of the ball so that they could be the ones to dictate the
 rhythm of the game’

It is extremely rare for intensified overt subjects of any of the types described above to occur in preverbal position. What is more, even where no overt intensifier is present, the postverbal position almost always implies a certain degree of intensification or contrastive focus.⁶ Preverbal overt subjects, on the other hand, have a fundamentally different function, that of identifying, individuating or disambiguating the agent/subject of the infinitival clause, which can, but need not, be different from the subject of the main clause (Schulte, 2007, pp. 178–187).

The fact that the overt subject of infinitival adverbial clauses introduced by *ao* cannot appear preverbally is due to fact that *ao* is, etymologically, a contraction of the preposition *a* and the masculine definite article *o*. Whilst this contraction has acquired the function of a preposition introducing temporal adverbial clauses and is no longer analysable as consisting of two separate functional elements when used in this construction, it continues to be a productively used contraction in all

5. In this paper, “contrastive focus” is understood as distinguishing the affected referent from all other potentially available referents in a given context; this can be achieved by the use of overt intensifiers, but also by means of prosodic stress and/or syntactic position. Contrastive focus is, in general terms, equivalent to “identificational focus”, as defined by Kiss (1998).

6. In some generative accounts (e.g., Madeira, 1994, p. 183) it is incorrectly claimed that the postverbal position is not available for the subject of infinitival adjunct clauses, thereby ignoring subjects bearing contrastive focus.

other contexts (cf. Torres Cacoullós & Bauman, in this volume, on the development of the preposition *para/pora* by chunking of prepositions *por* and *a*). Speakers therefore continue to identify the formal presence of the definite article in the [*ao* + infinitive]-construction, which in turn leads to the avoidance of (perceived) sequences of two definite articles, as in the ungrammatical example (8a).

- (8) a. **A-o o político faz-er uma visita a-o*
 at-the the politician make-INF a visit to-the
estado, disse [...]
 state said
 ‘When the politician paid a visit to the state, he said [...]’
- b. **A-o eu faz-er uma visita a-o*
 at-the PN.NOM.1SG make-INF a visit to-the
estado, disse [...]
 state said
 ‘When I paid a visit to the state, I said [...]’

In (8b) it can be seen that this restriction on preverbal subject NPs containing a definite article is analogically extended, triggering a generalised ban on preverbal subjects in the Portuguese [*ao* + infinitive]-construction. It should be noted that this analogical extension is not a necessary or obligatory one, as shown by the corresponding Spanish sentences in (9). Spanish also blocks preverbal full NP subjects in the [*al* + infinitive]-construction (9a), but licenses pronominal ones (9b), though the postverbal position is, nevertheless, more frequent even for pronominal subjects, as shown in Table 2 below.

- (9) a. **A-l el político hac-er una visita a-l*
 at-the the politician make-INF a visit to-the
estado, dijo [...]
 state said
 ‘When the politician paid a visit to the state, he said [...]’
- b. *A-l yo hac-er una visita a-l*
 at-the PN.NOM.1SG make-INF a visit to-the
estado, dije [...]
 state said
 ‘When I paid a visit to the state, I said [...]’

3.2.2 Portuguese in diachrony

As noted in Section 3.1.2, in Medieval Portuguese overt subjects frequently appear in pre-infinitival position since the very earliest preserved documents, whether

the infinitival clause functions as the subject, the object or an adjunct of the main clause. Although this is unusual from a comparative Romance perspective, it is in line with the general default word order and should, therefore, not come as a surprise; what it shows is that infinitival clauses with overt subjects were already fully integrated and syntactically unexceptional in Old Portuguese.

Beginning in the 16th century, a striking discrepancy between the position of pronominal and full NP subjects can be observed. In a corpus study of the complete works of Camões (16th century), Otto (1888, pp. 92–94) lists a total of 37 instances of personal infinitives with overt subjects, of which 34 occur in clauses introduced by the prepositions *de*, *por*, *para* or *sem*. Within these prepositional infinitive clauses, all seven instances of pronominal subjects appear preverbally, whilst only two of the 27 full NP subjects occupy that position.

- (10) [*para* *vós* *lhe* *faz-er-des* *tantos* *males*]
 [for PN.NOM.2PL PN.DAT.3SG do-INF-2PL so.many evils]
 ‘for you to do so many bad things to him’
 (Camões, *Sextina* 2, cited by Otto, 1888, p. 94)

- (11) [*sem* *fic-ar* *n-a* *alma* *a* *magoa* *d-o* *pecado*]
 [without remain-INF in-the soul the suffering of-the sin]
 ‘without the suffering of the sin remaining within the soul’
 (Camões, *Sonnet* 75, cited by Otto, 1888, p. 93)

This clear tendency for pronominal subjects to occur preverbally and full NP subjects postverbally, as shown in examples (10) and (11), is almost certainly related to the fact that, during this period, subjects in finite clauses most frequently appear in postverbal position (Eide, 2006, p. 46; in this volume, Section 5). Rather than being an SVO language, Classical Portuguese is a topic-verb-XP language (Eide, 2010, p. 147) in which the basic word order is highly dependent on information structure; subjects generally only precede the verb if they are also the topic of the sentence (Eide, in this volume). It appears that this pattern also affected the word order in OSI clauses; the preference for the preverbal position in the case of subject pronouns can be explained by the fact that their anaphoric nature makes it highly likely for them to refer to the topic in the majority of cases, whereas full NP subjects are more likely to refer to new information. In general terms, then, the word order in OSIs mirrors that of finite clauses in the 16th century, just as it does in Old Portuguese. However, there are differences regarding the degree to which this pattern becomes generalised. Whilst OSIs functioning as the subject or object of the main clause do occasionally have postverbal pronominal objects, as in example (12), the fact that, in adverbial infinitive clauses, all pronominal subjects, independently of their discourse function, appear preverbally, indicates that this originally pragmatically-based distribution pattern appears to have become

entrenched, pragmatically bleached and ultimately reanalysed as a syntactic rule. The explanation for this difference probably lies in the fact that it is more frequently necessary for an overt (pronominal) topical subject to appear in adverbial clauses for the purpose of disambiguation than in subject and object clauses, whose comparative closeness to the main verb generally make it unnecessary for a disambiguating, coreferential pronominal subject to appear overtly to indicate topic continuity. As a result, a greater proportion of preverbal pronominal subjects appear in adverbial clauses, triggering the generalisation of this pattern.

- (12) *Me atormenta então [ver eu que*
 PN.OBL.ISG torments then [see.INF PN.NOM.ISG that [...]]
 'It then torments me to see that [...]'
 (Camões, *Sonnet 270*, cited by Otto, 1888, p. 93)

The distribution pattern observed in Classical Portuguese is, however, short-lived. From the second half of the 17th century onwards, the postverbal position becomes more readily available for pronominal subjects, once again coinciding with a change in the word order of finite clauses, in this case particularly of subordinate clauses, which begin to exhibit a renewed preference for preverbal subject position at approximately the same time (Eide, 2006, p. 46). Gradually, SVO is re-established as the basic word order, leaving the postverbal subject slot available for specific discourse pragmatic functions.⁷ As observed by Martins (2011, p. 133), the rightmost constituent is pragmatically salient and has discourse/informational prominence in both Old and Modern Portuguese. In OSI clauses, it is contrastive focus that becomes increasingly associated with post-infinitival subjects, which can either be full NPs or subject pronouns.

The contrast between post-infinitival subjects with a specific, clearly defined discourse function (contrastive focus) on the one hand and pragmatically unmarked pre-infinitival subjects that are used for semantic disambiguation on the other is further reinforced by the fact that the personal (inflected) infinitive does not require the presence of an overt subject NP. Disambiguation of the subject's identity can be achieved by means of morphological inflection alone, bringing this type of OSI clause into line with other clause types found in pro-drop languages, like Portuguese, in which pragmatically unmarked subjects whose identity can be retrieved from the context are generally identified by means of person inflection on the verb. The person agreement morphology of the inflected infinitive has, thus, allowed it to become largely integrated into this general pattern.

7. It might, indeed, be speculated that the previous absence of post-infinitival pronominal subjects constituted a structural gap that lent itself to being exploited for a pragmatic purpose.

In summary, the diachronic development of the position of overt subjects in Portuguese prepositional infinitive clauses can be described as follows: in Old Portuguese, the subject tends to appear before the infinitive, as might be expected in an SVO language. In particular, the fact that the infinitive agrees with its subject in person and number shows that the relation between these two constituents is not fundamentally different from that between finite verbs and their subjects. During the classical period, the shift in basic word order causes full NPs, which often introduce new information, regularly to appear in post-infinitival position, whilst pronominal subjects, which are less likely to introduce new information, generally remain pre-infinitival, as seen in examples (10) and (11), respectively. This differential positioning of nouns and pronouns becomes entrenched and no longer necessarily depends on the actual discourse function in a specific context. The shift back to basic SVO word order in Modern Portuguese makes the – now prominent – post-infinitival subject slot available to mark contrastive focus, leading to the present-day functional split between pre- and post-infinitival subjects. During all three stages in the history of Portuguese, the rules governing word order in OSI clauses are closely linked to those of finite clauses.

3.3 The position of pronominal subjects of infinitives in Spanish adverbial clauses

3.3.1 *Modern Peninsular Spanish*

It is a well-known fact that infinitives are very often preceded by an overt subject pronoun in Caribbean varieties of Spanish (cf. Suñer, 1986; Morales de Walters, 1988). However, this dialectal peculiarity will not be discussed here, as it is better interpreted as part of a more general trend of these “pronoun-heavy” varieties towards using subject pronouns more frequently than in standard Spanish in a wider range of environments, such as second person contexts and *wh*-questions (cf. Lipski, 1977).

In standard Spanish, infinitives with overt subjects are relatively rare, and the general consensus in the literature is that the postverbal position is favoured, though it is acknowledged that pre-infinitival subjects are occasionally found in adverbial clauses (Gili Gaya, 1985, p. 189; Fernández Lagunilla, 1987, pp. 127, 132–133; Mensching, 2000, p. 25). A more detailed analysis reveals some striking differences within the area of adverbial clauses, as seen in Table 2, which shows the position of pronominal subjects in OSIs with the verbs *hacer* and *saber* in adverbial clauses introduced by different prepositions.⁸

8. The data used for compiling Table 2 are drawn from *Google* search results for sites within the .es-domain, i.e. located in Spain, on 15 April 2004. The figures presented here are *not* the approximate numbers of hits estimated by *Google*; each search result was verified individually.

Table 2. Position of pronominal subjects in OSIs

Preposition	Pre-infinitival pronominal subjects	Post-infinitival pronominal subjects
<i>sin</i> 'without'	68% (601)	32% (277)
<i>para</i> 'in order to'	46% (322)	54% (372)
<i>al</i> 'when'	28% (23)	72% (58)
<i>antes de</i> 'before'	21% (10)	79% (37)
<i>después de</i> 'after'	17% (3)	83% (15)

What the data in this table reveals is that the likelihood of a pronominal subject to be pre- or post-infinitival depends heavily on the prepositional element that introduces the infinitival clause, i.e. on the type of adverbial notion it expresses.

As in Portuguese, there is a clear link between the presence of overt intensifiers and the subject position, with a strong tendency to place intensified overt subjects postverbally. For instance, in the *para*-construction, around 96% of subject pronouns accompanied by intensifiers such as *mismo*, *también* or *solo* appear postverbally and only 4% preverbally. However, this distribution is less clear-cut with the other prepositional constructions. With *sin*, for instance, around 20% of overt intensifiers occur preverbally, as in example (13). A historically-based explanation for the special behaviour of the 'abessive' OSI construction with *sin* will be proposed in Section 3.3.2 below.

- (13) *Las parejas que no pueden divertirse solas van camino hacia*
 the couples that not can enjoy alone go path towards
el divorcio, muchas veces [sin ellos
 the divorce many times [without PN.NOM.M.PL
mismo-s sab-er-lo]
 themselves-PL know-INF-PN.ACC.3SG]

'Couples that can't have a good time on their own are on the road leading to divorce, without knowing it themselves'

Stylistic factors are also involved in the choice of position; for example, *sin saberlo ella* is generally considered less colloquial than the inverse. Mensching (2000, p. 26) remarks that pre-infinitival subjects are "characteristic for the spoken language", with most examples in literary texts appearing in direct speech.

A further factor that influences the choice between the pre- and post-infinitival subject position is the presence of an enclitic object pronoun on the infinitive. As shown in Table 3, with *para*, *antes de* and *después de*, the subject pronoun is significantly more likely to appear postverbally in the presence of an object clitic.

Table 3. Position of pronominal subjects in adverbial clauses

Adverbial clause	Pre-infin. pronom. subjects	Post-infin. Pronom. subjects	Adverbial clause
<i>para PN hacer/saber</i>	61% (282)	39% (179)	<i>para hacer/saber PN</i>
<i>para PN hacerlo/saberlo</i>	17% (40)	83% (193)	<i>para hacerlo/saberlo PN</i>
<i>antes de PN hacer/saber</i>	41% (9)	59% (13)	<i>antes de hacer/saber PN</i>
<i>antes de PN hacerlo/saberlo</i>	4% (1)	96% (24)	<i>antes de hacerlo/ saberlo PN</i>
<i>después de PN hacer/saber</i>	27% (3)	73% (8)	<i>después de hacer/ saber PN</i>
<i>después de PN hacerlo/ saberlo</i>	0% (0)	100% (7)	<i>después de hacerlo/ saberlo PN</i>
<i>al PN hacer/saber</i>	27% (16)	73% (44)	<i>al hacer/saber PN</i>
<i>al PN hacerlo/saberlo</i>	33% (7)	67% (14)	<i>al hacerlo/saberlo PN</i>
<i>sin PN hacer/saber</i>	53% (211)	47% (187)	<i>sin hacer/saber PN</i>
<i>sin PN hacerlo/saberlo</i>	81% (390)	19% (90)	<i>sin hacerlo/saberlo PN</i>

Faced with the data in Table 3, it has to be asked why the presence of an object pronoun generally appears to favour the post-infinitival position for the subject. Again, the explanation is related to the fact that contrastive focus is, statistically, associated with the post-infinitival position. If we consider that clitic pronouns are incompatible with the emphasis associated with contrastive focus, but that they nevertheless occupy one of the available argument slots, then the statistical likelihood of one of the other arguments (among them the subject) being focalised is greater in clauses containing a clitic. In clauses in which the direct object is a full NP rather than a clitic, on the other hand, the possibility of this object NP being the focalised element of the clause reduces the chances for one of the remaining arguments to be focalised. As a result, the relative proportion of focalised overt subjects in clauses containing a clitic is greater. The greater proportion of postverbal pronominal subjects in the presence of an object clitic can thus be understood as further evidence for the link between contrastive focus and post-infinitival subject position.

However, this pattern is not universal. The subject position in the construction with *al* shows very little sensitivity to the presence of clitics, and in the construction with *sin*, preverbal subject pronouns occur *more* frequently in the presence of a clitic pronoun. A look at the constructions' historical development casts some light on this split development in which the construction with *sin* is set apart from other prepositional infinitives.

3.3.2 Spanish in diachrony

Until the 18th century, the position of subject pronouns in infinitival clauses was less restricted than today (cf. Section 3.1.1). As Spanish is an SVO language, the unmarked position for a subject pronoun is left of the verb; *a priori*, there is no reason to expect this to be any different in OSI clauses, similar to the situation in Old and Modern Portuguese, as discussed in Section 3.2.2. According to the data for 16th-century Spanish provided by Keniston (1937, p. 550) and discussed by Ledgeway (2000, p. 292) (cf. Section 3.1.1.), there was a certain tendency for overt pronominal subjects to appear before the infinitive at the time, in a development somewhat similar to that in Portuguese during the same period. However, in Spanish this pattern did not become generalised or entrenched as it did in Portuguese, and postverbal pronominal as well as full NP subjects remain frequent, as in examples (14) and (15) (see Sitaridou, 2002, 2009). Overt subjects have always been optional in infinitival clauses, and the OSI construction was used for two specific purposes: (a) disambiguation, clarification or identification of the identity of the subject and (b) pragmatic functions such as emphasis, intensification and contrastive focus, as shown in the 16th-century examples in (14) and (15) respectively.

(14) Disambiguation

porque [*sin merec-er-lo yo*] [...], *me*
 because [without deserve-INF-PN.ACC.3SG PN.NOM.1SG] PN.OBL.1SG

proveyó y hizo merced de un beneficio en Toledo,
 supplied and made favour of a benefit in Toledo,

mi patria
 my hometown

‘because, without my deserving it, he provided for me and gave me support
 in Toledo, my hometown’ (Villegas, 1580–1603)

(15) Contrastive focus

Pienso que [en haz-er vosotros] de la misma manera, no
 think that [in do-INF PN.NOM.2PL] of the same manner not

erraréis
 will.err

‘I think that, by acting in the same way, you will not be wrong’ (Valdés, 1535)

In (14), the subject of the main clause is not coreferential with the subject of the infinitival clause in square brackets, and since the default assumption for infinitival clauses is subject coreference, the function of the overt pronoun *yo* is to override this and eliminate a possible misreading of the identity of the subject. In (15), on the other hand, the subject of *erraréis* is identical with that of the subordinate infinitive, *hazer*; there is, thus, no need for disambiguation in this case.

Instead, the overt subject pronoun is used for the purpose of contrastive focus, setting apart the addressees (*vosotros*) from those who already act in the way the speaker advises the addressees to act.

The two functions of the OSI were, thus, essentially equivalent to those in Portuguese, despite the fact that the infinitive cannot be inflected to agree with its subject in Spanish. However, this lack of morphological agreement on the infinitive always makes it necessary for the subject of an OSI to appear overtly as a noun or pronoun, which is what has caused the OSI construction to go down a different path in Spanish than in Portuguese. Since the overt presence of a subject pronoun in all other clause types is primarily associated with pragmatic functions such as emphasis and focus, the overt subjects in Spanish OSIs were increasingly reanalysed as having those pragmatic functions by default. The possibility of avoiding the appearance of disambiguating pronominal subjects in pragmatically unmarked contexts by using a finite clause, in which the identity of the subject can be disambiguated by means of verbal inflection alone, facilitated the OSI's increasingly specialised use for discourse pragmatic purposes, notably to mark contrastive focus. Furthermore, this type of discourse function is characteristically associated with prosodic stress, for which the overt presence of the focalised element is a prerequisite.

This development stands in stark contrast to that in Portuguese where the possibility of specifying the identity of the infinitive's subject by means of verbal inflection, without the need for an NP to be overtly present, brings the OSI more into line with finite clauses, in which person/number inflection alone is sufficient to identify a previously mentioned subject. As a result of this, no conflict arises between the presence of subject pronouns for the purpose of disambiguation and those used for contrastive focus. Just as in finite clauses, a disambiguating subject pronoun can, but need not, appear before the verb that agrees with it in person and number, which provides the possibility of avoiding pronominal subjects in cases where they might be reanalysed as focalising elements. In Spanish, on the other hand, the necessary overt (pronominal) subjects, even when intended to disambiguate the identity of the subject, are pragmatically salient and therefore prone to being reanalysed as containing a degree of contrastive focus, even if only in addition to their disambiguating function. This association between overt subjects in infinitival clauses on the one hand and contrastive focus on the other gradually causes a partial reanalysis of the construction itself, a process in which its discourse pragmatic use is increasingly turning into its central function.

This reanalysis, which is not yet completed, as shown by the fact that overt subjects can still be used to disambiguate the identity of the subject in adverbial clauses, has led to the present-day situation in which the pragmatically marked postverbal position is occupied by overt subjects more frequently than the unmarked preverbal one. The majority of these postverbal subjects do, indeed,

bear contrastive focus. However, even pragmatically unmarked, disambiguating subjects increasingly tend to occupy that position, which indicates that due to the high frequency of postverbal subjects, VS is becoming the default word order in Spanish OSI clauses. As a consequence, the postverbal subject position is no longer exclusively associated with its original discourse function, with the intriguing effect that the overall increase in discourse pragmaticity of the OSI construction triggers a process of ‘pragmatic bleaching’ of the post-infinitival subject position. This process has already been completed in OSIs functioning as subject or object of a sentence, where disambiguating as well as focalising subjects always appear post-infinitivally, as seen in examples (2a) and (3a) above, and it is all but finalised for full NP subjects even in adverbial infinitival clauses. As mentioned in Section 3.1.1, this tendency is already visible in the 16th century. With regards to pronominal subjects in adverbial OSIs, it was shown in Table 2 that clauses introduced by different prepositions are affected to a varying degree. Among these, the case of abessive OSI clauses with *sin* was shown to buck the trend and resist the general tendency towards postverbal subject position to a greater extent than other adverbial OSIs. In what follows, a more detailed analysis of abessive OSIs will reveal why this is the case.

Since the second half of the 15th century, a small set of verbs has been used very frequently with *sin* in OSI constructions, causing some particularly common [*sin* + infinitive + object clitic]-sequences to become virtually lexicalised, among them those shown in (16).

- (16) *sin saberlo* ‘unwittingly’
sin merecerlo ‘undeservedly’
sin pensarlo ‘unexpectedly’

As seen in example (14), the subject of these expressions is typically the experiencer of an event that has a different agent, necessitating the presence of a disambiguating subject pronoun, which is more likely to appear preverbally. Examples (17) and (18) are typical cases from Golden Age Spanish (16th–17th century).

- (17) *Me comenzaron [...] a inclinar-me a lo*
 PN.OBL.ISG began [...] to incline.INF-PN.OBL.ISG to that
que fue, [sin yo pensar-lo], mi perdición
 which was [without I think.INF-PN.AC.3SG] my downfall

‘They began to incline me to what was my downfall, even though I didn’t think of it as such’
 (Cervantes, 1605–1615, Chapter 28)

- (18) *Bien veo yo [...] que el cielo [...], [sin yo*
 well see I [...] that the sky [...] [without PN.NOM.ISG

merecer-lo *me* *envía* [...] *algunas personas* [...] [
 deserve.INF-PN.ACC.3SG] PN.OBL.1SG sent [...] some persons [...]]
 ‘I can well see that, without my deserving it, heaven sends me some people’
 (Cervantes, 1605–1615, Chapter 27)

The comparatively high frequency of these common expressions leads to the entrenchment of the sequence [*sin* + subject pronoun + infinitive + object clitic], explaining why the *sin*-construction uniquely prefers preverbal subject pronouns (cf. Table 2), especially in the presence of a clitic pronoun (cf. Table 3). Whilst there is some variability and this pattern is not always strictly adhered to, even with the typical verbs (cf. example (14)), it is, on the other hand, also extended to other verbs and cases of intensification and contrastive focus, as in (19), perhaps via a stage in which the preverbal subject fulfilled a syntactic and a discourse pragmatic function at the same time. This can be seen in (20), where the intensifier clearly indicates the presence of contrastive focus, whilst the overt subject is also necessary syntactically because it is not coreferential with the subject of the main clause.

(19) [*Sin yo mismo dar-me cuenta*], *quise sustituir*
 [without PN.NOM.1SG self give.INF-REFL account] wanted substitute
la falta de mi tierra [...] [
 the missing of my earth [...]]
 ‘Without noticing it myself, I wanted to substitute the absence of my
 homeland [...]’ (Lorda Vidal, 2012)

(20) *Mi voz* [...] *salía de mi boca* [*sin yo mismo*
 my voice [...] left from my mouth [without PN.NOM.1SG self
pod-er] *reconocer-la como mía*
 be.able-INF] recognise-PN.ACC.3SG as mine
 ‘My voice came out of my mouth without myself being able to recognise it
 as mine’ (García Monterrubio, 2011, p. 357)

The increasing incidence of preverbal subjects in the OSI construction with *sin* thus contrasts with the move away from pre-infinitival subjects in other environments. In both cases, the more frequent subject position becomes entrenched and gradually turns into the default option, regardless of whether its original function was that of contrastive focus or disambiguation.

4. Conclusions

Though Modern Spanish and Portuguese have very different patterns of positioning the overt subject in infinitival clauses in terms of frequency, the preverbal position being more common in Portuguese whilst Spanish shows a clear tendency

towards postverbal subjects, we can nevertheless assume a common origin for both patterns. The preverbal position was originally used for subject disambiguation and the postverbal one for pragmatic ends.

Portuguese moves towards restricting the function of postverbal subjects to intensification and contrastive focus. The use of the inflected infinitive, with or without a preverbal overt subject, brings the personal infinitive in line with the unmarked word order pattern of other clause types and with the optional pro-drop structure of Portuguese. In Spanish, on the other hand, the absence of an inflected infinitive blocks the full syntactic integration of OSIs, and they become increasingly pragmatically marked, which in turn causes the proportion of overt subjects occupying the pragmatically marked postverbal position to increase. As a result, this pattern becomes entrenched and gains further ground, making it available, and in some environments obligatory, for overt subjects that would previously have occupied the preverbal position.

It was further shown that specific, highly frequent patterns, such as the abessive OSI with *sin*, can become entrenched in their own right and buck the general trend or even move in the opposite direction. The fact that there is, more generally, a great degree of variation between the likelihood of subjects appearing pre- or post-infinitivally in different types of adverbial OSI clauses indicates that we are not dealing with a single syntactic phenomenon, and that the ‘overt subject infinitive’ is most appropriately analysed not as a single construction but as a cover term for a series of independent constructions, each of which has its own diachronic trajectory.

References

- Ambar, M. M. (1994). Aux to COMP and lexical restrictions on verb movement. In G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi & R. Zanuttini (Eds.), *Paths towards universal grammar. Studies in honor of Richard S. Kayne* (pp. 1–23). Washington DC: Georgetown University Press.
- Christen, S. (2001). Genitive positions in Baltic and Finnic languages. In Ö. Dahl & M. Koptjevskaja-Tamm (Eds.), *Circum-Baltic languages. Vol. 2: Grammar and typology* (pp. 499–520). Amsterdam: John Benjamins. doi:10.1075/slcs.55.09chr
- Docal, S. (2012). *Individualidad*. Original by R. G. Ingersoll. Retrieved from <http://www.oocities.org/pejll/inger_individual.htm>
- Eide, K. G. (2006). Word order structures and unaccusative verbs in Classical and Modern Portuguese. The reorganisation of information structure. (Unpublished doctoral dissertation). University of Oslo, Oslo, Norway.
- Eide, K. G. (2010). Prosody, information structure and word order changes in Portuguese. In G. Ferraresi & R. Lühr (Eds.), *Diachronic studies on information structure. Language acquisition and change* (pp. 87–116). Berlin: De Gruyter.

- Eide, K. G. (2018). Postverbal subjects of unaccusative verbs in the history of Portuguese. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morphology-syntax* (pp. 149–171). Amsterdam: John Benjamins. (in this volume) doi:10.1075/ihll.16.07eid
- Fernández Lagunilla, M. (1987). Los infinitivos con sujetos léxicos en español. In V. Demonte & M. Fernández Lagunilla, *Sintaxis de las lenguas románicas* (pp. 125–147). Madrid: El Arquero.
- Gili Gaya, S. (1985). *Curso superior de sintaxis española*. Barcelona: Bibliograf.
- Hernanz Carbó, M. L. (1982). *El infinitivo en español*. Bellaterra: Universidad Autónoma de Barcelona.
- Kenesei, I. (1994). Complementation in Finno-Ugric. Retrieved from <<http://www.nytud.hu/kenesei/publ/finnugr2.pdf>>
- Keniston, H. (1937). *The syntax of Castilian prose. The sixteenth century*. Chicago: University of Chicago Press.
- Kiss, K. E. (1998). Identificational focus versus information focus. *Language*, 74(2), 245–273. doi:10.1353/lan.1998.0211
- König, E. (2001). Intensifiers and reflexive pronouns. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (Eds.), *Language typology and language universals. An international handbook* (747–760). Berlin: De Gruyter.
- Kujawski, G. (2012). O escritor cyberpunk brasileiro finaliza sua nova obra *Vistarmada*. Retrieved from <<http://www.ufsm.br/alternet/zine/gkramos.html>>
- Ledgeway, A. (2000). *A Comparative Syntax of the Dialects of Southern Italy: A Minimalist Approach*. Oxford: Blackwell.
- Lipski, J. M. (1977). Preposed subjects in questions: Some considerations. *Hispania*, 60(1), 61–67. doi:10.2307/340393
- Lorda Vidal, I-C. (2012). La sombra del tiempo es alargada: exilios españoles en Londres. *Fronterad. Revista digital*. Retrieved from <<http://www.fronterad.com/?q=node%2F5897>>
- Madeira, Ana Maria. (1994). On the Portuguese inflected infinitive. *UCL Working Papers in Linguistics*, 6, 179–203.
- Martins, A. M. (2011). Scrambling and information focus in Old and Contemporary Portuguese. *Catalan Journal of Linguistics*, 10, 133–158.
- Mensching, G. (2000). *Infinitive constructions with specified subjects. A syntactic analysis of the Romance languages*. Oxford: Oxford University Press.
- Morales de Walters, A. (1988). Infinitivo con sujeto expreso en el español de Puerto Rico. In R. M. Hammond & M. C. Resnick (Eds.), *Studies in Caribbean Spanish dialectology* (pp. 85–96). Washington, DC: Georgetown University Press.
- Otto, R. (1888). *Der portugiesische Infinitiv bei Camões*. Erlangen: Junge & Sohn.
- Posner, R. (1996). *The Romance languages*. Cambridge: Cambridge University Press.
- Raposo, E. (1987). Case theory and Infl-to-Comp: The inflected infinitive in European Portuguese. *Linguistic Inquiry*, 18, 85–109.
- Rouveret, A. (1980). Sur la notion de proposition finie: Gouvernement et inversion. *Langages*, 60, 75–107. doi:10.3406/lgge.1980.1863
- Schulte, K. (2007). *Prepositional infinitives in Romance: A usage-based approach to syntactic change*. Berne: Peter Lang.
- Sitaridou, I. (2002). *The synchrony and diachrony of Romance infinitives with nominative subjects*. (Unpublished doctoral dissertation). University of Manchester, Manchester, UK.
- Sitaridou, I. (2009). On the emergence of personal infinitives in the history of Spanish. *Diachronica*, 26(1), 36–64. doi:10.1075/dia.26.1.02sit

- Skydsgaard, S. (1977). *La combinatoria sintáctica del infinitivo español*, Vol. 1. Madrid: Castalia.
- Suñer, M. (1986). Lexical subjects of infinitives in Caribbean Spanish. In O. Jaeggli & C. Silva-Corvalán (Eds.), *Studies in Romance linguistics* (pp. 189–203). Dordrecht: Foris.
- Torrego, E. (1998). Nominative subjects and pro-drop INFL. *Syntax*, 1(2), 206–219.
doi:10.1111/1467-9612.00008
- Torres Cacoullós, R., & Bauman, J. (2018). Allative to purposive grammaticalization: A quantitative story of Spanish *para*. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 195–221). Amsterdam: John Benjamins. (in this volume) doi:10.1075/iHLL.16.9tor
- Yoon, J. H-S., & Bonet-Farran, N. (1991). The ambivalent nature of Spanish infinitives. In D. Wanner & D. A. Kibbee (Eds.), *New analyses in Romance languages: Selected papers from the 18th Linguistic Symposium on Romance Languages, Urbana-Champaign, 1988* (pp. 353–379). Amsterdam: John Benjamins. doi:10.1075/cilt.69.25yoo

Corpus

- Cervantes, M. de. (1605–1615). *El ingenioso hidalgo Don Quijote de la Mancha*. F. Rico (Ed.). Madrid: Instituto Cervantes, 1998.
- García Monterrubio, S. (2011). *La granja de la font*. Madrid: Visión.
- Ministério da Ciência e da Tecnologia (MCT) and Público. (2000). *Corpus de Extractos de Textos Electrónicos MCT/Público*. Retrieved from <<http://www.linguateca.pt/cetempublico>>
- Valdés, J. de. (1535). *Diálogo de la lengua*. J. M. Lope Blanch (Ed.). Madrid: Castalia, 1976.
- Villegas, Alonso de. (1580–1603). *Fructus sanctorum y quinta parte del Flos sanctorum*. J. Aragüés Aldaz (Ed.). València: Universitat de València, 1998.

Allative to purposive grammaticalisation

A quantitative story of Spanish *para*

Rena Torres Cacoullos & Joseph Bauman

The Pennsylvania State University

The Spanish preposition *para* arises from fusion of older *por* followed by *a*, via usage-based chunking (Bybee, 2010) associated with the frequency of the sequence. At an early stage, semantic compositionality involving an independent meaning contribution from the preposition *a* is discernible in allative uses with a destination that is a person. A general decline of allative uses of *para* with a nominal complement, for both person and place destinations, ensues after the 14th century. In a second change beginning in the 18th century, *para* replaces *por* with infinitive complements to become the majority purposive variant. Thus, allative-to-purposive grammaticalisation (Heine & Kuteva, 2002) occurs through changes both in *para*'s distribution across contexts and in its variation with the older preposition.

Keywords: chunking; compositionality; frequency; grammaticalization; preposition; allative; purposive; Spanish

1. Introduction

The story of Spanish *para* is fascinating because it allows observation of how a new preposition arises, evolves and spreads in the linguistic system. *Para* (via its 12th–13th century precursor *pora*) arises from the fusion of two earlier prepositions, *por* (or its Latin antecedent(s) *per* and *pro*) and *a(d)*. A corresponding preposition developed in Portuguese (*para*) and Catalán (*per a*), but not in Romance languages beyond the Iberian Peninsula (cf. Luraghi, 2010, p. 69). Germanic languages, too, e.g., with English *for* or Dutch *voor*, may be said to have one adposition corresponding to Spanish's two, *por* and *para*.

From the earliest Spanish examples, *para* appears in two contexts. One is with NP complements, with a meaning of motion to or toward a location (1a) or with abstract meanings, such as benefactive (1b). A second early context for the new

preposition is with infinitive complements as an expression of purpose, in variation with older *por* (2b) vs. (2a).

- (1) *para* + NP
 - a. Spatial (allative)

*Adelino **para** San Pero, olas dueñas estan*
 ‘He headed to San Pedro where the ladies are staying’
 (12th c., *Cid*, 1392)¹
 - b. Abstract (benefactive)

*Yo adobare conducho **para** mi et para mis vassallos*
 ‘I will prepare food **for** me and for my vassals’
 (12th c., *Cid*, 249)
- (2) *por* vs. *para* + Infinitive
 - a. *noe tendio la mano **por** tomar la*
 ‘Noah reached out his hand **to** take it [the dove]’
 (13th c., *GEI*, 14r)
 - b. *le escogio entre todos **para** dezir le su uolu<n>tad*
 ‘He chose him among all others **to** tell him His will’
 (13th c., *GEI*, 11v)

After presenting the database (Section 2), we begin by tracing the origins of *para* (Section 3). The fusion of the two prepositions, *por* and *a*, is tied to frequent co-occurrence as predicted by the hypothesis that with repetition, sequential experiences become ‘chunked’ into new units (Bybee, 2010, p. 34). Changing distribution across contexts of occurrence indicates that *para* evolved through loss of compositionality, manifested in an early [*para* + NP_{specific human}] construction (Section 4). A first overarching change that we observe is that spatial uses steeply decline as a proportion of *para* + NP occurrences between 14th and 15th century texts (Section 5). A second change, observed between the 17th and 18th centuries, is that the newer preposition, *para*, becomes more associated with the purposive infinitive to the detriment of the older one, *por* (Section 6).

2. Data

The corpus compiled for this study comprises 17 texts, beginning with *El cantar de mio Cid* (1140–1207) and drawing on two prose texts for each subsequent century up to the 20th. From the 16th century onward, one of these is a peninsular text and the other is a chronologically corresponding sample from the *Documentos lingüísticos de la Nueva España* (New Spain/Mexico). Texts were selected based on the availability of reliable digital versions (see Company Company, 2006, pp. xxiv–xxvii on the corpus for the *Sintaxis histórica de la lengua española*).

1. Within parentheses following examples we indicate the century, the abbreviated name of the text (see Corpus section) and the verse, chapter, page or folio number for the edition listed.

For each century, approximately 1,000 tokens of *por* and *para* with a nominal/adverbial or infinitive complement were extracted (that is, not extracted were tokens of *por* or *para* followed by the conjunction *que* and a finite verb). Note that counts for *para* include tokens of the form *póra* in the 12th and 13th centuries. Information on the sampling procedure and exclusions is given in Torres Cacoullós & Bauman (2014, pp. 391–393).

Table 1 shows the texts and token counts. The editions are listed in the Corpus section. The last column in Table 1 gives the frequency of *para* relative to *por* by century.² As we will see (Section 6), the frequency increase of the newer preposition relative to the older one from the 18th century onward has occurred disproportionately in infinitive constructions.

Table 1. Texts by century, token counts (*para* and *por*) and frequency of *para* relative to *por*

Cent	Text	N	% <i>para</i> *
12th	<i>Cid</i>	370	17%
13th	<i>Calila</i> , pp. 91–181	406	17%
	<i>GEI</i> , pp. 5–122	563	
14th	<i>Zifar</i> , pp. 9–110	500	24%
	<i>Lucanor</i> , odd numbered <i>exempla</i>	572	
15th	<i>Corbacho</i> , pp. 67–104 (I, 1–17), p. 145–183 (II, 1–8)	511	21%
	<i>Celestina</i> , pp. 67–214	556	
16th	<i>LT</i>	348	30%
	<i>DLNE</i> 1535–1569, pp. 109–161 (docs. 17–36)	364	
17th	<i>Quijote II</i> , odd numbered chapters between 1 and 27	489	23%
	<i>DLNE</i> 1609–1640, pp. 240–347 (docs. 79–129)	495	
18th	<i>CN/Sí</i>	339	32%
	<i>DLNE</i> 1790–1810, pp. 611–709 (docs. 258–307)	495	
19th	<i>Regenta</i> , chapters 16, 19, 22, 25	433	33%
	<i>Bandidos</i> , pp. 27–278 (chapters 1–29)	503	
20th	<i>Madrid</i> , pp. 87–290 (transcripts 5–16)	630	32%
	<i>México</i> , pp. 11–172 (transcripts 1–13)	491	

* % *para* in each century combines data from the two texts. 12th and 13th century *para* counts include instances of *póra*. Counts of *póra* and *para* are, respectively, 63 and 1 in the *Cid*, 78 and 1 in *GEI*, 1 and 83 in *Calila*.

2. The figures in Riiho (1979, p. 234), which include cases of *por/para* + *que*, also indicate increasing frequency of *para* relative to *por*: 12th 8%, 13th 12%, 14th 13%, 15th 9%, 16th 22%, 17th 23%, 18th 33%, 19th 29%, 20th 26%.

3. Chunking: *para* < *pōra* < *por* + *a*

3.1 Frequency and fusion

Most scholars agree that *para* arises from the fusion or coalescence (Haspelmath, 2011) of *por* and *a*(*d*). In seeking evidence for this, we were delighted to find 12th–13th century examples in which *por* and *a* appear together in non-agglutinated form. A search of the 600,000-word digital version of the *GEI* edited by Kasten, Nitti and Jonxis-Henkemans (1997) yields 30 tokens of *por* + *a* separated by a space (not agglutinated). Example (3) illustrates variation between the non-agglutinated and the agglutinated form in this text, in a near-identical context.

- (3) *por* + *al* vs. *poral* variation in *GEI*
diz q<ue> algun bien es por al om<n>e en no<n>
seer pecador. & maguer que non es sa<n>cto. como diz q<ue> es otrossi algo
poral om<n>e del qui non puede seer Rey (13th c., *GEI*, 57v)

The majority of the tokens of *por* + *a* separated by a space in the *GEI* occur with a noun as the object of the preposition and with the definite article fused with *a* in *al*, as in the first line in (3) above and in the examples in (4). Most tokens of *por* + *a* separated by a space in the 13th century *La Fazienda de Ultramar* are also cases with *al* (Dave McDougall, p.c.). It seems, then, that the construction which most resisted the fusion of *por* and *a* is with the masculine definite article *el*, perhaps due to counteraction from the stronger contraction of *el* to the preposition *a*. Supporting this account is the persistence of the contracted form *poral*, examples of which are found until the 17th century (Riiho, 1979, p. 236).

- (4) Non-agglutinated *por* + *al* + Noun
- a. Spatial/allative
uinieron se de su uagar por al mont synay
 ‘They returned from their travels at Mount Sinai’ (13th c., *GEI*, 176v)
fueron se amos el mont ariba por al logar del sacrificio
 ‘They both went up the mountain to the place of the sacrifice’
 (13th c., *GEI*, 63r)
 - b. Temporal
et lo q<ue> uos sob<r>a `re condesad lo por al sabb<ad>o
 ‘And what is left over, save for the Sabbath’ (13th c., *GEI*, 171r)
 - c. Abstract
si era sana ell animalia & q<u>a `l deuie seer por al sacrificio.
 ‘If the animal were healthy and as it should be for the sacrifice’
 (13th c., *GEI*, 173r)

Nevertheless, there are examples of *por + a* in other contexts, such as with a pronoun (5a) and a bare noun (5b) in *La Fazienda de Ultramar*. We find a non-agglutinated example of *por a* preceding an infinitive in the *GEI* (5c).³ Thus, although tokens with a space between *por* and *a* are a small minority, there are enough to suggest that scribes utilised the non-agglutinated combination with some regularity in 13th century texts.

- (5) Non-agglutinated *por + a* in other contexts
- a. *prēdre auos por ami.*
por mio pueblo.
 ‘I will take you [literally: to/for me] as my own people’
 (*Fazienda* 13vA25, Ex.6:7, McDougall, p. c.)
- b. [...] *accarrear lenna e agua*
por accaffa del ciador.
 ‘Bring firewood and water to/for (the) house of the creator’
 (*Fazienda* 29vA24, Josh.9:23, McDougall, p. c.)
- c. *por a yr a Egipto a ueer su fio*
 ‘To go to Egypt to see his son’
 (13th c., *GEI*, 108v)

It has been stated that two-preposition combinations are common in Spanish (e.g., Alonso & Henríquez Ureña, 1959, § 230; Keniston, 1937, § 41.41). Besides *por + a*, we also find sequences of *por* with prepositions other than *a*, as with the examples in (6) (as well as with adverbs, for example, *por arriba* ‘above’, *por dentro* ‘inside’, *por encima* ‘top’). Some combinations are more coherent than others; for example, the coherence of *por de* in (6a) would appear to be greater than that of the *por con* sequence in (6b), which is the fortuitous result of the versification (in *por con ellos lidiar*, the Preposition + Pronoun combination *con ellos* ‘with them’ interrupts the *por + Infinitive* combination *por lidiar* ‘(in order) to combat’), while *por* and *en* in (6c) seem to be semantically overlapping.⁴

- (6) *por + preposition other than a*
- a. *la reyna Casiope presciaua se por fermosa & aun por de buenas*
costumbr<e>s
 ‘Queen Cassiopeia was esteemed for (being) beautiful and for (being) of good conduct’
 (13th c., *GEII*, 216v)

3. We are grateful to Dave McDougall for providing these examples. The agglutinated form *ami* (*a + mi*) (5a) appears in the *Fazienda de Ultramar* (see the palaeographic transcription from Arbesú (2011)).

4. We thank an anonymous reviewer for suggesting this formulation.

- b. *Grandes son los poderes por con ellos lidiar*
 ‘Great is their strength to combat with them’ (12th c., *Cid*, 669)
- c. *comendaro<n> ell Jmperio de Costa<n>tinopla a aquel dicho Rey*
Joh<a>n. por en toda su uida.
 ‘They entrusted the empire of Constantinople to said King John for
 (in) all his life’ (13th c., *Estoria de España II*, 285r)

There has been much discussion on whether the original first element in *para* is specifically *pro* (Corominas, 1980–1983, see *para*; Menéndez Pidal, 1940, § 129; Meyer-Lübke, 1890–1906, III.168) or *per* (Hanssen, 1945, § 726). The general meaning of *pro* is ‘in front of’ (with the ablative) while that of *per* is ‘through, over’ (with the accusative) (Allen and Greenough, 1931, pp. 134–135). Alternatively, *para* could be seen as having derived from the confluence of both *pro* and *per* with *ad*. Viewing the former as conveying a sense of orientation and the latter a trajectory through space, Melis (1992, pp. 71–76) suggests that in combination with *ad*, contributing its own sense of approaching a limit, the prepositions yield *para*’s basic meaning of a trajectory oriented toward a limit. While consensus has not been reached, examples of combinations of *a(d)* with both *per* and *pro* are attested in Medieval Latin documents in Spain (9th–12th centuries), with spatial as well as abstract meanings, illustrated in (7a) and (7b), respectively (Melis, 1992, p. 77). These seem similar to the 13th century Spanish uses of *por* + *a* illustrated in the set of examples in (4) above.

- (7) *per/pro* + *ad* + NP in Medieval Latin (Spain; cited by Melis, 1992, p. 77)
- a. Spatial
vadit de Villanova per ad Sancta Maria
 ‘[he] goes from Villanova towards Santa Maria’ (*Fueros*)
- b. *jlla karera que descure pro ad Uiginagio*
 ‘The path that runs back to Viginagio’ (*Crestomatía*)
- c. Abstract
pectet V solidos, medios per ad opus de illo senior
 ‘(he) pays five solidos, half for the lord’s work’ (*Fueros*)
Dono [...] pro ad illo ganato [...] una baccariza
 ‘I give [...] for the herd [...] a corral/enclosure’ (*Cartulario*)

A requirement for the creation of a new unit is frequent co-occurrence of its erstwhile component parts. From the perspective of a usage-based approach to grammatical structure, a mechanism for the creation of constituent structure is the “chunking” of a sequence of morphemes that results from frequent repetition of the sequence (Bybee, 2010, p. 34; cf. Moyna, in this volume). “Items that are used together fuse together” per Bybee’s Linear Fusion Hypothesis (Bybee, 2002, p. 112). This may occur even in the absence of semantic coherence, as with the

contraction of auxiliary *will* to the first person pronoun *I* rather than to the following verb (Bybee, 2010, pp. 136–138). Bybee (2002) shows that *I will* or *I'll* is twice as frequent as *will be*, the most frequent *will* + Verb combination (in the Switchboard corpus).

In accordance with the hypothesis of chunking, the fused unit (> *pora/para*) would follow from the repeated sequence of the two prepositions *por* + *a*. The prediction is that for the sequence *por* + *a* to result in a fused unit (where *por* in combination with other prepositions does not), the co-occurrence of *por* + *a* must be more frequent than other combinations. The data support this prediction, as illustrated in Figure 1 below.

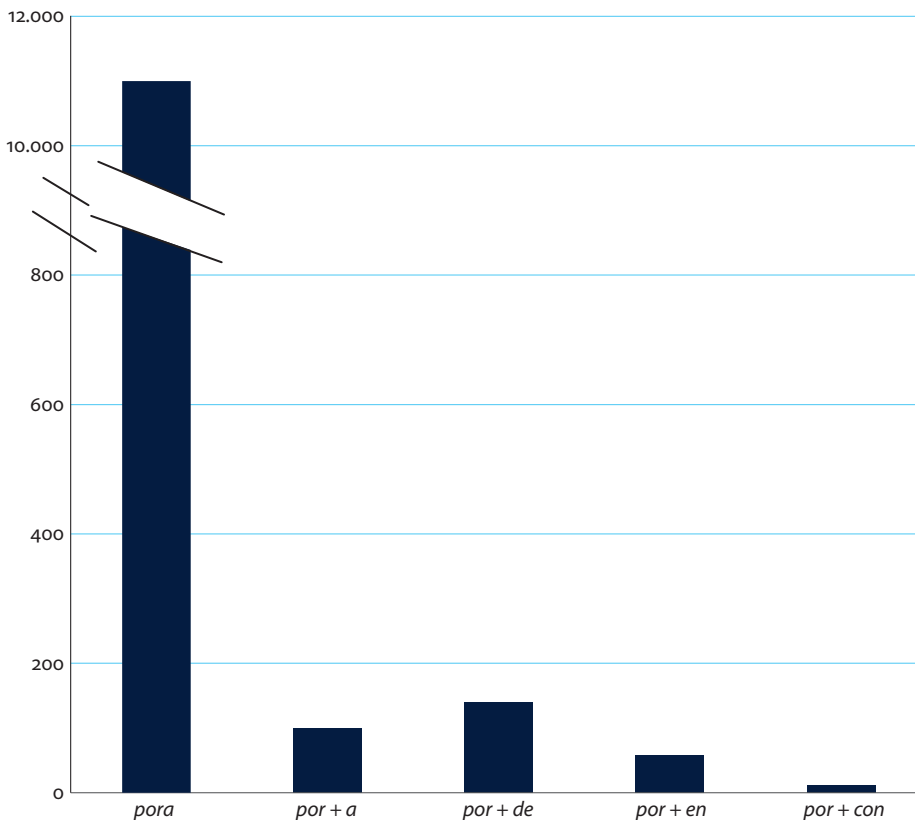


Figure 1. Token frequency *por* + preposition (*Corpus del español* counts for 1200s)

Figure 1 shows token frequency counts for 13th century texts drawn from the online corpus of Mark Davies. At first glance, the *por* + *a* sequence may appear to not be much more frequent than the other *por* + preposition sequences (and indeed, less frequent than *por* + *de*). However, for the purposes of tabulating

co-occurrence of *por + a* vs. *por +* another preposition, we count together *pora* and *por + a* (i.e., all cases as in (1), (2b), (3), (4) and (5)) given variation in this period between the agglutinated and non-agglutinated forms (example (3), above) and the lack of consistency in transcriptions (see below). When combined with the already orthographically fused *pora* – more than 10,000 tokens, compared with fewer than 200 cases of *por + de* – it is clear that the *por + a* sequence is far more common than any of the other combinations.

3.2 Precursors, variant forms and palaeographic abbreviations

Earlier Spanish texts are marked by variation between the modern *para* and the older form *pora*.⁵ We find 63 tokens of *pora* in the 12th or 13th century *Cid* (8a), compared to one token of *para* (9a), denoting a spatial relation (a destination). In the analysed portion of the 13th century *GEI*, we count 83 tokens of *pora* (8b) versus only one of *para* (9b).⁶ In the 14th century *Zifar*, however, there are no tokens of *pora*, and in *Lucanor* there is just one (Exemplo XLII, p. 219), the latest in our corpus.

- (8) Earlier form *pora*
- a. *Vansse pora San Pero*
'They go to San Pedro' (12th c., *Cid*, 294)
 - b. *estas serien despues pora comer pora ell omne*
'These would be then for men to eat' (13th c., *GEI*, 12v)
- (9) In variation with *para*
- a. *para Calatayuch quanto puede se va*
'As soon as he can he goes to Calatayud' (12th c., *Cid*, 775)
 - b. *fuel aparta<n>do toda uia para si e a los suyos.*
'He divided it [the land] for himself and his own' (13th c., *GEI*, 5r)⁷

This corroborates the statement that *pora* is the form that “is typically found” (our translation) in the 12th and 13th centuries (Corominas, 1980–1983, see *para*). According to Riiho’s (1979, p. 232) token counts, the relative frequency

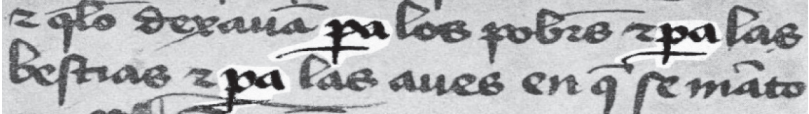
5. Other related forms have also been noted: *pera* (Espinosa, 2010, p. 216), *perad* (García de Diego, 1951, p. 128), *par* (Cuervo, 1886–1994, see *para*; García de Diego, 1951, p. 217). The phonetic and/or phonological processes argued to account for the transition from these forms to the modern *para* are discussed by García de Diego (1951, p. 128), Meyer-Lübke (1890–1906, III.518), and Brea (1985, p. 167).

6. Although they are theoretically contemporary texts, there are only 15th century copies of *Calila* available to us, while the *GEI* is the only part of the complete *GE* that is dated at the end; we therefore know that it is from the late 13th century.

7. In example (9b) *para* and *a* appear in the same context.

of *pōra* with respect to *para* decreases precipitously from 83% to 15% from the first to the second half of the 13th century, and from there to 2% in the first half of the 14th until disappearing completely by the second half of that century. We cannot know for sure if this is due to a fairly rapid sound change or if it was the graphical conventions used to represent the forms that abruptly changed (on *pōra* and other forms preceding *para*, see Torres Cacoullós & Bauman, 2014, pp. 403–414).

As concerns graphical conventions, we are confronted with another issue that lends support to the assumption that *pōra* and *para*, together with non-agglutinated *por + a* (Section 3.1, above), are best viewed as a single entity albeit with different orthographic and possibly phonetic manifestations. Given that at least three variants were being used simultaneously during the 13th century, it is no surprise that scribes would sometimes resolve this variation by using a fourth abbreviated form. In (10), the preposition appears three times as a conventionalised, ambiguous abbreviation without the interior graphemes, displaying only the initial *p* and the final *a*.

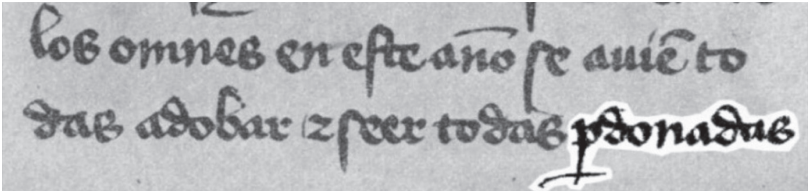
(10) 

et q̄lo dexauā pa los pobrē et pa las bestias et pa las aues [...]

‘and that he left it for the poor and for the beasts and for the birds [...]

(Solalinde, 1930, p. 424)

The abbreviation with a horizontal bar either crossing or connecting with the descender (the vertical line) of the *p* was already a convention in the transcription of Latin *per* and *pro* (Cappelli, 1990, p. 257). In Cappelli’s dictionary of abbreviations, the horizontal bar does not cross the descender in the abbreviation for *pro*, but merely connects with its left side. With this in mind, the form of the *p* that is encountered in Medieval Spanish manuscripts, as in (10), more closely resembles the abbreviations used for *per*. Indeed, we find another example of the abbreviation, this time involving the adjective *perdonadas* (the adjectival form of the verb *perdonar* ‘to pardon’, which is unequivocally composed of the prefix *per-* and the root *donar* ‘to give’), shown in (11).

(11) 

los omnes en este año se auie to

das adobar et seer todas pdonadas

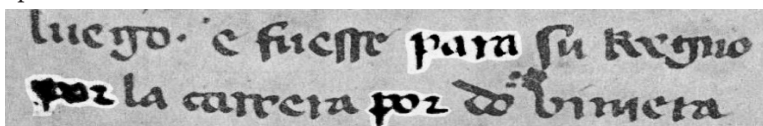
‘the men in that year all had to dress and be pardoned’ (Solalinde, 1930, p. 424)

This clear abbreviation of *per-* using a *p* with a crossbar does not mean we must conclude that the examples of *pa* necessarily represent *pera*, *pora*, or some specific variant. Rather, all that is certain is that this abbreviation was used to represent a vowel-consonant combination (or perhaps consonant-vowel, in light of the frequent transposition of the *r* in Old Spanish (Corominas, 1980–1983, see *por*)), as with the combinations of a vowel and /r/ in *pora*, *para*, and *per-*.

Still, it is possible that the transcriptions from which we have extracted our data do not faithfully represent the original forms used by medieval scribes. More than 100 years ago Hanssen (1911) mentioned the “little trust that scribes, and at times also editors, deserve” (our translation) and 25 years ago Melis (1992, p. 82, n. 24) similarly alluded to the occasional lack of exact correspondence between original documents and modern transcriptions. Nevertheless, in reproductions of the original manuscripts of the *GE* that Solalinde (1930) includes, unequivocal tokens of *para* already appear (12). The new preposition indicates a spatial relationship in (12a) and an abstract relationship in (12b). The example in (12b) also provides another illustration of the occasional disagreement between manuscripts and modern transcriptions. Solalinde (1930, p. 287) transcribes this token as *pora* in his edition although it appears as *para* in the reproduction of the original manuscript on the immediately preceding page (1930, p. 287, folio VII).

(12) *para* in 13th century original manuscript

a. Spatial

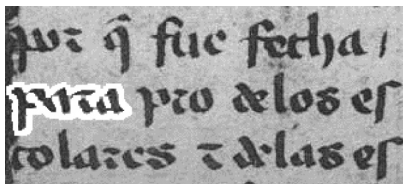


luego e fuesse para su regno por la carrera por dō viniera

‘then and he went to his kingdom by the road by which he had come’

(Solalinde, 1930, p. 488)

b. Abstract



por q̄ fue fecha para pro delos escolares [...]

‘by which it was made for the good of the scholars’ (Solalinde, 1930, p. 287)

It is clear, then, that for some time the orthographic variants *por a*, *pora*, and *para* were used contemporaneously. Rather than a linear, consecutive evolution (*per/pro + ad > pora > para*), the coexistence of these forms in the same text indicates

that the process of the fusion of the two (groups of) structural elements was gradual and proceeded via variation between forms, as is expected with processes of change.

To summarise this section, despite the vexing problem of manuscript editions and graphic conventions, there are orthographic indications that the combination of the prepositions *por* and *a* as independent elements may have persisted into early (13th century) Spanish texts. Furthermore, consistent with the usage-based hypothesis of chunking with repetition, the *por* (+) *a* sequence is of high frequency. This pair of facts provides evidence for the view that the new preposition *para* is an outcome of the fusion of two prepositions.

Now, where the orthographic and phonetic sequence of events must remain nebulous, there is a solid source of evidence for the origins of *para* from the semantic functions of the new preposition, as we will see below.

4. Compositionality of *para*, and its loss

Additional evidence that *pora/para* arises from the combination of two separate prepositions, one of which was *a*, may be adduced from early compositionality of meaning.⁸ As put forward in Bybee (2010, pp. 44–45; see also Croft & Cruse, 2004, pp. 250–253; Langacker, 1987, p. 292), compositionality is a semantic measure having to do with transparency of meaning, and refers to the degree to which the meaning of the whole is predictable from the meaning of the component parts. In this sense, *hopeful* is more compositional than *awful* (the meaning contribution of *hope* to the derived word is more discernible than that of *awe*).⁹

Here we implement this semantic measure by operationalising semantic compositionality of a linguistic form in terms of the contexts of use of the linguistic form. In particular, we consider the kinds of object NPs with which *para* co-occurs.

8. Compositional meaning in *para* has been claimed for early examples such as *Çercar quiere a Valençia pora christianos la dar* (*Cid*, 1191) ‘he seeks to besiege Valencia to give it to the Christians’, in which “la forma *a* tiene su función propia, distinta y separada de la de *por*” ‘the form *a* has its own function, separate and distinct from that of *por*’ (our translation) (Riiho, 1979, p. 99; see also Riiho, 1979, p. 124).

9. In contrast with the semantic criterion of compositionality, *analysability* is a morphosyntactic parameter, referring to the degree to which the internal structure and individual parts are recognisable; for example, while *pull strings* is not fully compositional because it has a metaphorical meaning it is analysable in that speakers are assumed to recognise an individual verb and its noun complement (Bybee, 2010, p. 45).

In early texts we find examples in which the subject of the motion verb modified by the *para* phrase reaches their destination. In (13), for example, *para* indicates the terminal point of the subject's movement, a use that we associate with the preposition *a* in present-day Spanish. The subject not only goes 'toward' the *posada* but actually reaches it, as verified by the fact that he speaks to the person there.

- (13) [motion verb + *para* + NP] = to a location
E el pleteo con ellos e fuese para la posada e dixole su muger commo auia
pleteado con los marineros
 'And he Settled with them and went to the inn and told his wife how he had
 Settled with the seamen' (14th c., *Zifar*, 87)¹⁰

The endpoint of the motion may even be a person. In the examples in (14), the object NP of *para* has a human referent; again, in present-day Spanish we would expect not *para* in this context but the preposition *a*.

- (14) [motion verb + *para* + NP_{specific human}]
 a. *Venimos nós para ti que nos consejes*
 'We came to you so that you may advise us' (13th c., *Calila*, 144)
 b. *E el moço se fue para su padre, e dixo la respuesta*
 'And the boy went to his father, and told him the answer' (14th c., *Zifar*, 21)
 c. *y assi me fuy para mi amo, que esperandome estaua.*
 'And thus I went to my master, who was waiting for me' (16th c., *LT*, 1)

In this motion-verb *para* construction with a destination that is a person, the human referent of the object appears as a personal pronoun ($N = 20$) or definite full NP (in which the determiner is a definite article ($N = 16$), possessive ($N = 13$), or demonstrative).¹¹ Furthermore, this NP with a human referent overwhelmingly appears in singular number (92%, 46/50, of 13th–14th c. tokens). From these nominal features we may infer that the referent is specific and individuated. This is precisely the kind of referent we would expect the preposition *a* to co-occur with, in accordance with its use as a dative marker (indirect objects tend to be human) and progressively also as an accusative marker for direct objects referring to persons (on the latter, see Company Company, 2002).

10. There is no dative *a* preceding *su muger* in example (13) in the edition of *Zifar* we used.

11. In *Calila*, we count as human the personified animals.

Figure 2 shows the proportions of spatial uses of *para* + NP with prepositional objects that are persons, as in (13) above, as opposed to places (as in (12)). Object NPs with a human referent as the endpoint of motion constitute a full third, 35% (50/144), of spatial instances of *para* in 13th–14th century texts. A handful of cases (4/15) are still found in 15th–16th texts. But the [motion verb + *para* + NP_{specific human}] construction disappears from the corpus after the 16th c. (0/51).

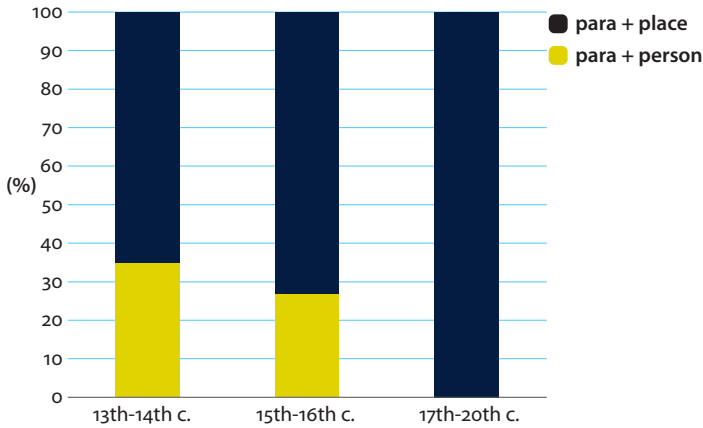


Figure 2. Spatial uses of *para* + NP, proportion of object NPs designating persons vs. places

We submit that this early [motion verb + *para* + NP_{specific human}] construction indicates that the semantic contribution of the preposition *a* was still discernible in the new preposition, and thus a degree of compositionality in its beginnings (cf. Riiko, 1979, p. 99). Subsequently, object NPs with a human referent as the endpoint of motion disappear with *para*, which indicates that *a* no longer makes an independent meaning contribution.

To summarise this section, the semantic measure of compositionality of meaning is implemented to adduce evidence that *para* arises from a combination of two separate prepositions, one of which was *a*. Operationalising compositionality of *para* in terms of the kinds of NP objects with which it co-occurs in spatial relations, we observe a robust [motion verb + *para* + NP_{specific human}] construction in early texts. Quantitative change in the contexts of use of *para*, with the decline and disappearance of destinations that are persons, indicates erosion of the semantic contribution of *a* as it is absorbed into the new preposition, and thus loss of compositionality of the whole.

5. Early allative use: Distributions of the preposition across its contexts of use

The uses of a preposition or the kinds of relations it signals may be classified into spatial, temporal or abstract (or, notional) (e.g., Pottier, 1968, p. 25; Delbeque, 1996, p. 252). Classifying thus tokens of *para* with other than an infinitive complement, usually an NP, we find that, while the proportion of temporal uses (e.g., *para mañana*, *para las diez*) has remained steady at approximately 5% to 10% throughout the centuries, there is a notable shift in the proportion of spatial uses (Torres Cacoullós & Bauman, 2014, p. 421). In this section we provide evidence for early allative use and its decline.

Table 2 shows the most frequent motion verbs modified by a *para* phrase among tokens of the preposition classified as spatial (rather than temporal or abstract). The grouping of the centuries into three time periods, 12th–14th, 15th–16th and 17th–20th, emerged once we considered each century separately. In the 12th–14th century data, three verbs of directional motion—*ir(se)* ‘go’, *venir(se)* ‘come’, *tornar(se)* ‘return’—constitute 80% (132/164) of all spatial instances of *para* + NP, as shown in last column of Table 2. The preponderance of directional motion verbs is consonant with an allative meaning of movement to or towards a location.

It is relevant that middle-marked *irse* is somewhat more frequent than *ir* (at a ratio of 1.3 to 1, or 56 to 44 tokens). This skewing contrasts with that observed in the same period for spatial uses of *por*, with which *irse* is ten times less frequent than the unmarked *ir* (at a ratio of 1 to 10, or 3 to 31 tokens). The co-occurrence of *para* with *se*-marked motion verbs may be viewed as further evidence for the early use of *para* as an allative, if the middle marker has telic meaning, focusing on the moment of change for the experiencer of the motion event (Maldonado, 1999, Ch. 6; Torres Cacoullós & Schwenter, 2008).

Table 2. Most frequent motion verbs modified by *para* and their proportion of the preposition’s spatial uses

Century	<i>Ir</i>	<i>Irse</i>	<i>Tornar(se)</i>	<i>Venir(se)</i>	%
12th–14th	44	56	13	19	80% (132/164)
15th–16th	6	0	0	1	29% (7/24)
17th–20th	7	8	0	4	40% (19/47)

Frequent co-occurrence with *ir(se)* ‘go’, which alone constitutes close to two-thirds of spatial instances in the earliest period, is important for the evolution of *para*, in light of a generalisation from typological studies that “in the presence of ‘go’-verbs,

allatives frequently take on a purposive reading” (Rice & Kabata, 2007, p. 459). An example would be German *zu*, whose “allative use [...] evolves into a purposive one” (Lehmann, 2002, p. 6). Another example of allative > purpose grammaticalisation is Lezgian *-z*, which evolves from a direction marker as a nominal suffix (15a) to a purposive marker as a verbal suffix (15b) (Haspelmath, 1993, p. 89, p. 156, cited in Heine & Kuteva, 2002, p. 39).

(15) Allative-to-purposive grammaticalisation in Lezgian

a. *Zun medinstitutdi- z fi- da.*
I:ABS medical:school-DAT go- FUT
‘I’ll go to medical school’

b. *I irid stxa čpi- n juldaš-ri-q^h galaz*
this seven brother selves-GEN friend-PL-POSESESS with
q̄uğwa-z fe-na.
play-INF go-AOR
‘These seven brothers went to play with their friends’

What is meant by “evolves from ... into”? What happens with *para* is certainly not an abrupt or even linear replacement of one use by another. From the earliest texts, allative and purposive uses coexist, as illustrated in (16) (see also (14a)).

(16) Coexistent allative (16a) and purposive (16b)

a. *fueron-se para la ribera de la mar*
go.PFV.3PL-REFL to ART.DEF.F.SG shore of the sea

b. *para se y- r*
to REFL go-INF

‘They went to the shore to depart’ (14th c., *Zifar*, 89)

Rather, “semantic change [...] should be manifested in changing distribution and co-occurrence patterns” (Torres Cacoullos & Schwenter, 2007, p. 357). One measure is the distribution of *para* across its contexts of occurrence. The line marked with diamonds in Figure 3 shows the proportion of spatial uses of *para* by century (not counting occurrences with an infinitive complement). First, it is evident that even in the earliest texts, *para* expresses spatial relations in only (approximately) half (43%–52%) of its occurrences followed by an NP (or adverb), that is, we do not find an initial period in which *para*, or its precursor *pora*, exclusively or even mostly had a spatial sense. Second, after the 14th century, during which *para* definitively displaces *pora* in Spanish texts (Riiho, 1979, p. 232), the proportion of spatial uses is no greater than 10%.

To confirm that the decline of spatial uses is not a mere accident of genre or topic, the line marked with squares in Figure 3 shows the corresponding distribution of *por* tokens. Naturally, in those texts in which the protagonist, for example,

the *Cid*, *Lazarillo* or *Don Quijote*, moves around, rates of spatial uses are higher than in texts such as the *Corbacho*, which is largely a sermon. Nevertheless, despite such fluctuations we can discern the stability in the overall proportion of spatial uses of *por*, which ranges from 10% to 30%. This contrasts with the patent drop in the proportion of spatial uses of *para*.

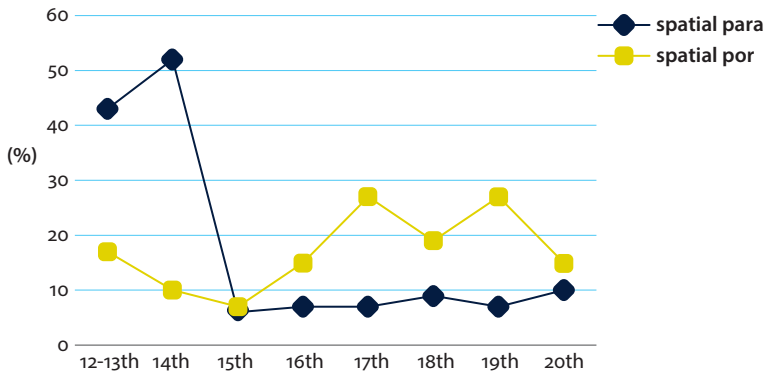


Figure 3. Proportion of spatial uses of *para*, compared with spatial uses of *por* (not counting infinitive complements for either)¹²

To summarise this section: in its beginnings *para* may be characterised as an allative, based on co-occurrence with directional motion verbs, often *se*-(middle-) marked, and spatial uses constituting approximately half of all tokens of *para* (not counting infinitive complements). However, not only do specific human destinations disappear, as we saw in the previous section, but there is an overall decrease in the proportion of spatial uses, which drop after the 14th century, as *para* with a nominal complement is used mostly to express non-spatial, abstract relations (such as benefactive, e.g., (1b)).

A second major quantitative change in co-occurrence patterns concerns infinitive complements. Here it is not distribution across contexts of occurrence that provides an enlightening measure, but variation with respect to the older preposition, *por*.

12. For Figure 3, Ns *para*: 12th–13th 75/173; 14th 89/170; 15th 7/120; 16th 8/115; 17th 9/122; 18th 12/134; 19th 9/122; 20th 26/261. Ns *por*: 12th–13th 164/959; 14th 71/724; 15th 51/713; 16th 55/370; 17th 160/587; 18th 92/487; 19th 157/575; 20th 85/579.

6. Rising purposive: Variation between *para* and *por* with infinitives

Table 3 provides two frequency measures of *para* + Infinitive. The first considers distribution of the form across contexts of occurrence, namely the proportion of tokens of *para* with an infinitive complement, shown in the first pair of columns. No clear rising trend is discernible. What is revealing is variation with respect to the older preposition in this context. Shown in the second pair of columns is the frequency of *para* relative to *por* with an infinitive complement. Here we see a reversal in the relative frequency of *por* and *para*, which are evenly distributed (~50%) in the [+ Infinitive] context until the 17th century, after which the rate of *para* increases, reaching 85% in the 20th century.

Figure 4 displays the increasing frequency of the newer preposition relative to the older one overall (Table 1), in the line marked with squares, juxtaposed to its increasing relative frequency in the particular context of infinitives (Table 3, second pair of columns), in the line marked with diamonds. It is clear that the frequency increase of *para* with respect to *por* has occurred disproportionately precisely in [+ Infinitive] constructions.

Table 3. Frequency of *para* + Infinitive (a) as a proportion of all tokens of *para*; (b) relative to *por* + Infinitive

Century	% <i>para</i> : proportion <i>para</i> with an infinitive	N	% <i>para</i> : relative to <i>por</i> with an infinitive	N
12th–13th	34%	77/226	41%	77/187
14th	33%	84/258	53%	84/159
15th	45%	100/221	46%	100/219
16th	43%	92/212	46%	92/201
17th	48%	111/232	41%	111/272
18th	47%	127/268	68%	127/187
19th	60%	183/307	79%	183/232
20th	42%	200/471	85%	200/235

Now, it is generally understood that, in present-day Spanish, infinitive constructions with *por* mean something different from those with *para*, the former expressing cause (or reason), the latter purpose. For example, in (17), with *por*, the subject felt guilty because he married off someone, whereas in (18), with *para*, the subject needed money in order to retrieve his clothing.

However, it is apparently common cross-linguistically for the same form to have both a purpose and a reason sense (among other polysemies of allatives),

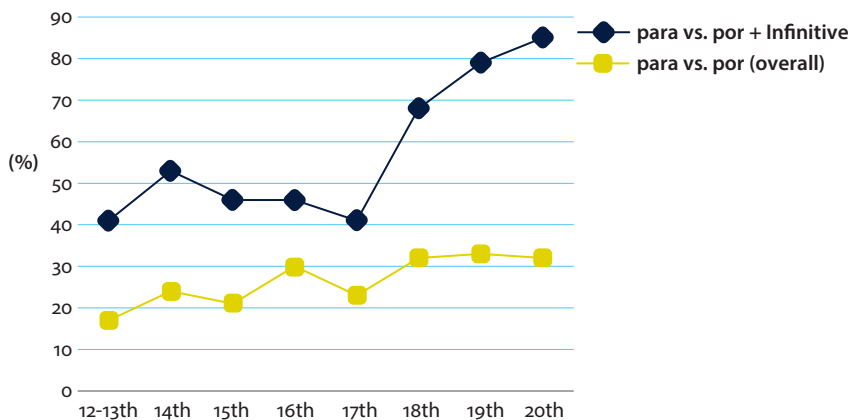


Figure 4. Increasing relative frequency of *para* vs. *por* overall (Table 1) and more striking increase in relative frequency of [*para* + Infinitive] vs. [*por* + Infinitive] (Table 3)

as with English *for* (Rice & Kabata, 2007, p. 455). As we will see just below, for a good part of the history of Spanish, [verb + *por* + Infinitive] was used in a purposive sense. A purposive *por* is illustrated in (19), from the 15th century *Celestina*, where we might well expect to find *para* today: the meaning here is that she wants to trick him not because she is rich, but in order to become so.

- (17) [verb + *por* + Infinitive] = cause (reason)
Se creía [...] culpable por haber casado a Tules
 ‘He thought himself [...] guilty for having married off Tules’
 (19th c., *Bandidos*, 1)
- (18) [verb + *para* + Infinitive] = purpose
necesitaba diez pesos para sacar su ropa empeñada
 ‘He needed money **in order to** retrieve his clothing that had been pawned’
 (19th c., *Bandidos*, 1)
- (19) [verb + *por* + Infinitive] = purpose
También quiere a mí engañar como a mi amo por ser rica
 ‘She wants to trick me as well as my master **in order to** become rich’
 (15th c., *Celestina*, 5)

As a replicable measure, we operationalise the purpose sense in terms of the temporal reference of the situation. We count as purposives those cases—of either [*por* + Infinitive] or [*para* + Infinitive]—in which the situation referred to by the infinitive is posterior to that of the main (finite) verb (see examples in (18)–(19)).

Beginning with the earliest appearance of the new preposition there has been variation between *por* and *pora/para* in the purposive infinitive construction.

Figure 5 shows the relative frequency of [*por* Infinitive_{Purposive}] and [*para* + Infinitive_{Purposive}] in a sample of the texts. In the 15th century *Celestina* and 17th century *Quijote*, purposive infinitive complements are evenly distributed between the two prepositions (50% (35/70) and 46% (28/61) for *para* in the two texts, respectively), but in the 19th century *Regenta* (Spain) and *Bandidos* (Mexico) the relative frequency of *para* is up to 90% (111/123).

From this we may conclude that *para* has generalised in the purposive infinitive construction, while *por* has greatly receded. In Modern Spanish, [*por* + Infinitive_{Purposive}] persists in particular constructions, such as to modify aspectual verbs *acabar*, *comenzar*, *concluir* (for example, *Comenzaron por apostar una botella* [*Bandidos*, 1.177]), and with negated infinitives (Torres Cacoulllos & Bauman, 2014, pp. 440–443).¹³

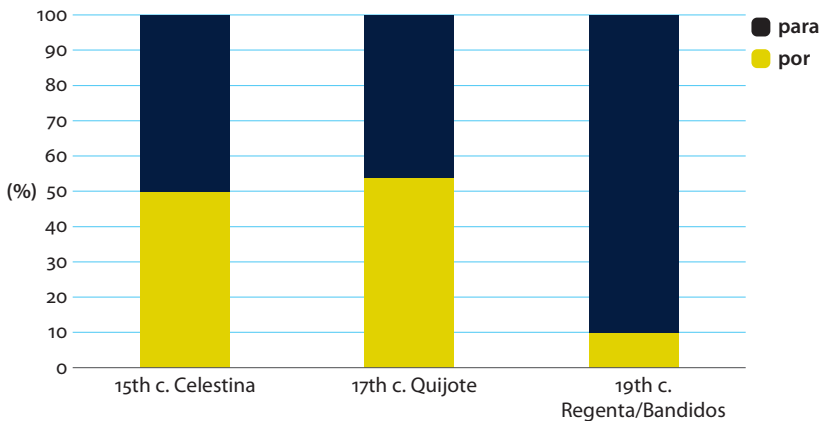


Figure 5. Rate of *para* relative to *por* with a purposive infinitive complement ([*para* + Infinitive_{Purposive}] vs. [*por* + Infinitive_{Purposive}])¹⁴

Nonetheless, it might be objected that [*por* + Infinitive_{Purposive}] and [*para* + Infinitive_{Purposive}] have never meant the same thing and therefore there really has been no replacement of *por* by *para*. In other words, even as ‘purposives’ there may

13. The drop in [*por* + Infinitive] between the 17th and 18th centuries may be parallel to the substitution of [*por* + *que* + Subjunctive] by [*para* + *que* + Subjunctive] purposive clauses (Bolinger, 1945, p. 16). Contributing to the drop in [*por* + Infinitive] may have been a rise in the use of [*porque* + Finite verb]-clauses to express cause (Torres Cacoulllos & Bauman, 2014, p. 440).

14. For Figure 5, Ns *por/para* + Infinitive: *Celestina* 70, *Quijote* 61, *Regenta/Bandidos* 123.

be a meaning difference, such that purposive infinitive clauses with *por* express a nuance of “underlying motive or incentive” (Bolinger, 1945, p. 20) that is absent from *para*. According to this view, the increased relative frequency of *para* would reflect a sociocultural change rather than genuine linguistic change. As Bolinger put it,

The intent of *por* to designate underlying motive or incentive also explains its relative infrequency as compared with *para*. Modern writing in general, and especially modern fiction, is far more objective than that of any preceding epoch. Where Cervantes was not averse to revealing the underlying motives and inner feelings of his characters, a modern writer would feel that he had no right to act the part of omniscience, and would prefer to let his readers guess at those internal secrets. (Bolinger, 1945, p. 20)

How can we determine objectively whether the rise of *para* in Figure 3 and Figure 4 reveals real linguistic change and not altered literary conventions? Evidence for a meaning difference between *por* and *para* again would be quantitative differences in co-occurrence patterns. Tabulating the co-occurring contextual elements of two apparently alternative forms offers a replicable way of characterising the semantic distinction between them.

The working hypothesis of the variationist method is that “within a given locus of variability, or *variable context*, [... the] competing variants will occur at greater or lesser rates depending on the features that constitute the context” (Poplack, 2001, p. 405). We define purposive infinitive complements (i.e., infinitives designating a situation temporally posterior to that of the finite verb) as a variable context and use this as a “heuristic device” to probe meaning differences (Poplack & Tagliamonte, 2006, p. 78). If linguistic forms mean something different, they should be preferred in different sub-contexts, in other words, they should each occur at greater than average rates in certain subcontexts (see also Aaron & Torres Cacoullous, 2005, p. 615; Kapatsinski, 2009, pp. 160–161).

We thus examine purposive infinitive constructions in a particular sub-context, that in which the subject NP has a human referent. Our reasoning is that subject NPs with a human referent are congruent with a meaning of intention (“underlying motive or incentive”, in Bolinger’s (1945, p. 20) terms), whereas inanimate subjects are incapable of having motives (cf. Melis, 1997, p. 112). The prediction, then, is that if purposive [*por* + Infinitive_{PURPOSIVE}] conveys a meaning of ‘underlying motive’ or intention absent from [*para* + Infinitive_{PURPOSIVE}], we expect *por* rates relative to *para* to be higher in the sub-context of human subjects (as in (20) and conversely *para* rates should be higher in the sub-context of inanimate subjects (as in (21)).

- (20) [human subject + verb + *por/para*+ Infinitive_{Purposive}]
- querría quedar por aliviar tu cuytado*
'I would like to stay so as to alleviate your sorrow'
(15th c., *Celestina*, 2)
 - quiso turbarme por oirme decir otras docientas patochadas*
'He tried to upset me so as to hear me say another two hundred follies'
(17th c., *Quijote* II, 7)
 - quería vencerla, para no padecer tanto*
'He wanted to overcome it [his ambition], so as to avoid suffering so much'
(19th c., *Regenta* II, 16)
- (21) [inanimate subject + verb + *por/para*+ Infinitive_{Purposive}]
- para qué es la fortuna favorable y próspera sino para servir a la honrra*
'For what is fortune favourable and propitious if not to serve honour'
(15th c., *Celestina*, 2)
 - toda la natura se remiró por la hazer perfecta*
'All of nature exerted itself to make her [Melibea] perfect'
(15th c., *Celestina*, 1)
 - el esfuerzo de tantos y tantos miserables servía para minarle el terreno*
'The effort of so many wretches served to undermine his ground'
(19th c., *Regenta*, II, 22)

In the examples in (21), with inanimate subjects for the finite (main) verb (fortune, nature, the effort), the instance with *por* (21b) is no more compatible with the expression of a nuance of intention in the purposive clause than those with *para* (21a), (21c). On the other hand, in the examples in (20), with volitional human subjects, there would appear to be no justification for considering the instances with *por*, in (20a) and (20b), as conveying more of a sense of underlying motive than the instance with *para*, in (20c) (besides the argument that *por* itself has such a meaning, which would be a circular argument).

Figure 6 shows the rate of *para* relative to *por* with a purposive infinitive complement when the prepositional phrase modifies a verb with a human subject, in the same sample of texts as for Figure 5. Comparing Figure 6 with Figure 5, we see that in the 15th century *Celestina* and 17th century *Quijote* the relative frequency of [*para* + Infinitive_{Purposive}] in the context of a human subject (at 46% (26/57) and 42% (22/53), respectively) is somewhat lower than the overall rate with a purposive infinitive in the corresponding text (Figure 5). This is because the rate of [*para* + Infinitive_{Purposive}] relative to [*por* + Infinitive_{Purposive}] is lower with a human subject than with an inanimate subject (at 76% (16/21), combining numbers from the two texts because of low token counts, $p < .03$ in Fisher's exact test). Thus, we can say that in these 15th and 17th century texts, human subjects

favor [*por* + Infinitive_{Purposive}], as predicted by the hypothesis of a (immutable) meaning difference between the two forms.

Nevertheless, in the 19th century texts, there is an inversion between the two prepositions and now *para* predominates (with a relative frequency of 89% (102/114)), even in the context of a human subject, our operationalisation of the expression of “underlying motive”. Furthermore, the ratio of purposive infinitives with a human vs. an inanimate subject for the main verb (counting tokens of both *por* and *para*) has not decreased, as would be expected if modern authors were eschewing “revealing the underlying motives” of their characters (Bolinger, 1945, p. 20) (at 5-to-1 in the *Celestina* and *Quijote* (110 to 21) and 13-to-1 in *Regenta* and *Bandidos* (114 to 9)). Thus, what we have evidence for is genuine linguistic change. Whereas in the 15th and 17th century in the sub-context of a human subject (to whom readers may attribute intentions) we tend to have *por*, as in (20a) and (20b), in the 19th century in the same sub-context we tend to have *para*, as in (20c). The change accords with the variationist hypothesis of “neutralization in discourse” of meaning differences (Sankoff, 1988, p. 153).

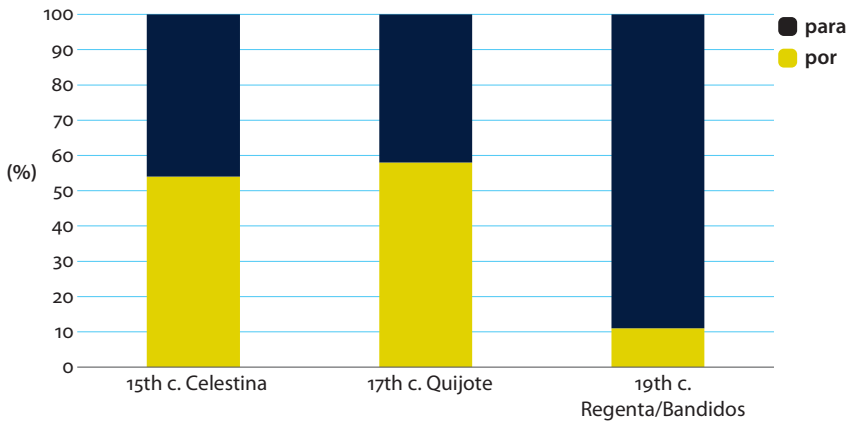


Figure 6. Rate of *para* relative to *por* with a purposive infinitive complement modifying a verb with a human subject

In summary, we first observed a reversal in the relative frequency of *por* and *para* after the 17th century in infinitive constructions overall (Figure 4). We then observed a parallel reversal more particularly with a purposive infinitive complement (counting as purposives those that are temporally posterior to the main verb) (Figure 5). Finally, we verified the same reversal in the rate of *para* relative to *por* in the environment of purposive infinitive constructions with a human subject in the main verb (Figure 6). The conclusion is that *para* has generalised as a purposive

infinitive marker.¹⁵ Together with the decline of allative uses after the 14th century and the displacement of *por* from infinitive constructions, this is taken as evidence for the allative > purposive grammaticalisation path.

7. Conclusion: A path and process of grammaticalisation

By drawing on grammaticalisation and variation theory we have offered a rich understanding of the rise of a new preposition. Grammaticalisation theory puts forward specifiable paths of change, such that particular lexical source meanings develop into particular grammatical meanings following similar paths (for example, verbs meaning ‘want’ participate in constructions that become futures), and processes of change, of which pertinent here is chunking and ensuing loss of compositionality.

The evidence adduced here for the evolutionary *path* followed by Spanish *para* has been two sets of changes. First is the decline of spatial allative uses of *para* + NP (after the 14th century), which includes the disappearance of destinations that are persons—NP objects with human referents (after the 16th century). The second is the increasing rate of *para* relative to *por* within purposive infinitive constructions (beginning in the 18th century). This pair of changes constitutes a quantitative demonstration, the first as far as we are aware, of the hypothesised cross-linguistic grammaticalisation path whereby allatives develop into purposives (Heine & Kuteva, 2002, p. 39).

As to the *process* of change by which the new preposition arises, we have shown loss of compositionality. *Para* begins as a sequence of two independent prepositions that, with frequent co-occurrence, are fused into a single grammatical unit. This process is known as chunking, which may be viewed as a domain-general cognitive process (cf. Bybee, 2010, p. 34). Operationalising semantic compositionality by observing prepositional objects designating destinations that are persons rather than locations, we showed an early [motion verb + *para* + NP_{specific human}] construction, evidence for an independent meaning contribution of *a*, which is subsequently lost. The story of *para* is thus served well by grammaticalisation theory, with its predictions about paths of change, here allative-to-purposive, and processes of change, here loss of compositionality.

15. See Torres Cacoullós and Bauman (2014, pp. 440–444) on the retention of [*por* + Infinitive] in particular constructions.

Change is (manifested in) alterations in quantitative co-occurrence patterns of the incoming linguistic form. What we have shown here is that two kinds of quantitative pattern are important. One, perhaps the more familiar relative frequency measure in historical linguistics, is the distribution of a linguistic form across its contexts of occurrence, seen here in the decline of spatial uses of *para*. But, as we saw with the rise of *para* and demise of *por* with purposive infinitives, crucial are the patterns of variation of the newer form with respect to the older one with which it comes to compete in a variable context.

References

- Aaron, J. E., & Torres Cacoullós, R. (2005). Quantitative measures of subjectification: A variationist study of Spanish *salir(se)*. *Cognitive Linguistics*, 16(4), 607–633.
doi:10.1515/cogl.2005.16.4.607
- Allen, J. H., & Greenough, J. B. (1931). *New Latin grammar for schools and colleges founded on comparative grammar*. J. B. Greenough, A. A. Howard, G. L. Kittredge & B. L. D'ooze (Eds.). New York, NY: Ginn and Company.
- Alonso, A. & Henríquez Ureña, P. (1959). *Gramática castellana, Curso 2o*. Buenos Aires: Losada.
- Bolinger, D. L. (1945). Purpose with *por* and *para*. *Modern Language Journal*, 28(1), 15–22.
doi:10.2307/317177
- Brea, M. (1985). Las preposiciones, del latín a las lenguas románicas. *Verba*, 12, 147–182.
- Bybee, J. L. (2002). Sequentiality as the basis of constituent structure. In T. Givón & B. F. Malle (Eds.), *The evolution of language out of pre-language* (pp. 109–134). Amsterdam: John Benjamins. doi:10.1075/tsl.53.07byb
- Bybee, J. L. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511750526
- Cappelli, A. (1990). *Lexicon Abbreviaturarum: Dizionario di abbreviature latine ed italiane*. Trento: Ulrico Hoepli Editore.
- Company Company, C. (2002). Grammaticalization and category weakness. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 201–215). Amsterdam: John Benjamins. doi:10.1075/tsl.49.14com
- Company Company, C. (2006). Introducción. In C. Company Company (Ed.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal* (Vol. 1; pp. xi–xxxiii). Ciudad de México: Fondo de Cultura Económica/Universidad Nacional Autónoma de México.
- Corominas, J. (1980–1991). *Diccionario crítico-etimológico castellano e hispánico*. Madrid: Gredos.
- Cuervo, R. J. (1994). *Diccionario de construcción y régimen de la lengua castellana*. Bogotá: Instituto Caro y Cuervo.
- Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511803864
- Delbecque, N. (1996). Towards a cognitive account of the use of the prepositions *por* and *para* in Spanish. In R. Dirven, R. Langacker & J. Taylor (Eds.), *Cognitive linguistics in the red-woods: The expansion of a new paradigm in linguistics* (pp. 249–318). Berlin: Mouton de Gruyter. doi:10.1515/9783110811421.249

- Espinosa Elorza, R. M. (2010). *Procesos de formación y cambio en las llamadas “palabras gramaticales”*. Logroño: Cilengua.
- García de Diego, V. (1951). *Gramática histórica española*. Madrid: Gredos.
- Hanssen, F. (1911). Cuestiones de gramática. Observaciones sobre la preposición PARA. *Bulletin Hispanique*, 13, 40–43. doi:10.3406/hispa.1911.1688
- Hanssen, F. (1945). *Gramática histórica de la lengua castellana*. Buenos Aires: El Ateneo.
- Haspelmath, M. (1993). *A grammar of Lezgian*. Berlin: Mouton de Gruyter. doi:10.1515/9783110884210
- Haspelmath, M. (2011). The gradual coalescence into “words” in grammaticalization. In B. Heine & H. Narrog (eds.), *The Oxford handbook of grammaticalization* (pp. 342–355). Oxford: Oxford University Press.
- Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511613463
- Kapatsinski, V. (2009). Adversative conjunction choice in Russian: Semantic and syntactic influences on lexical selection. *Language Variation and Change*, 21(2), 157–173. doi:10.1017/S0954394509990068
- Keniston, H. (1937). *The syntax of Castilian prose. The sixteenth century*. Chicago, University of Chicago Press.
- Langacker, R. (1987). *Foundations of cognitive grammar: Theoretical prerequisites*, Vol. 1. Stanford, CA: Stanford University Press.
- Lehmann, Ch. (2002). New reflections on grammaticalization and lexicalization. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 1–18). Amsterdam: John Benjamins. doi:10.1075/tsl.49.03leh
- Luraghi, S. (2010). Adverbial phrases. In P. Baldi & P. Cuzzolin (Eds.), *New perspectives on historical Latin syntax* (pp. 19–107). Berlin: Mouton de Gruyter.
- Maldonado, R. (1999). *A media voz: Problemas conceptuales del clítico se*. Ciudad de México: Universidad Nacional Autónoma de México.
- Melis, Ch. (1992). La preposición *para* del español: Un acercamiento a sus orígenes. In R. Barriga Villanueva & J. García Fajardo (Eds.), *Reflexiones lingüísticas y literarias, volumen I: Lingüística* (pp. 69–86). Ciudad de México: El Colegio de México.
- Melis, Ch. (1997). Sobre los inicios de la función ‘final’ de *para (que)*. In C. Company (Ed.), *Cambios diacrónicos en español* (pp. 99–121). Ciudad de México: Universidad Nacional Autónoma de México.
- Menéndez Pidal, R. (1940). *Manual de gramática histórica española*. Madrid: Espasa-Calpe.
- Meyer-Lübke, W. (1890–1906). *Grammaire des langues romanes. III. Syntaxe*. Paris: Welter.
- Moyna, M. I. (2018). The history of concatenative compounds in Spanish. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 47–74). Amsterdam: John Benjamins. (in this volume) doi:10.1075/iHLL.16.03moy
- Poplack, S. (2001). Variability, frequency and productivity in the irrealis domain of French. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 405–428). Amsterdam: John Benjamins. doi:10.1075/tsl.45.20pop
- Poplack, S., & Tagliamonte, S. (1996). Nothing in context: Variation, grammaticization and past time marking in Nigerian Pidgin English. In P. Baker & A. Sycia (Eds.), *Changing meanings, changing functions: Papers relating to grammaticalization in contact languages* (pp. 71–94). London: Westminster University Press.
- Pottier, B. (1968). *Lingüística moderna y filología hispánica*. Madrid: Gredos.

- Rice, S., & Kabata, K. (2007). Crosslinguistic grammaticalization patterns of the ALLATIVE. *Linguistic Typology*, 11, 451–514. doi:10.1515/LINGTY.2007.031
- Riiho, T. (1979). *Por y para. Estudio sobre los orígenes y la evolución de una oposición prepositiva iberorrománica*. Helsinki: Societas Scientiarum Fennica.
- Sankoff, D. (1988). Sociolinguistics and syntactic variation. In F. Newmeyer (Ed.), *Linguistics: The Cambridge survey, 4: Language: The socio-cultural context* (pp. 140–161). Cambridge: Cambridge University Press.
- Torres Cacoullós, R., & Bauman, J. (2014). Las preposiciones *por*, *pora* y *para*. In C. Company Company (Ed.), *Sintaxis histórica de la lengua española. Tercera parte: Preposiciones, adverbios y conjunciones. Relaciones interoracionales* (pp. 1479–1564). Ciudad de México: Fondo de Cultura Económica y Universidad Nacional Autónoma de México.
- Torres Cacoullós, R., & Schwenter, S. A. (2005–2007). Towards an operational notion of subjectification. In R. T. Cover & Y. Kim (Eds.), *Proceedings of the 31st Annual Meeting of the Berkeley Linguistics Society: General session and parasession on prosodic variation and change* (pp. 347–358). Berkeley, CA: Berkeley Linguistics Society.
- Torres Cacoullós, R., & Schwenter, S. A. (2008). Constructions and pragmatics: Variable middle marking in Spanish *subir(se)* ‘go up’ and *bajar(se)* ‘go down’. *Journal of Pragmatics*, 40, 1455–1477. doi:10.1016/j.pragma.2008.01.005

Corpus

- Bandidos* = Manuel Payno, *Los bandidos de Río Frío*. In *Obras completas*, M. Sol (Ed.), Ciudad de México: Consejo Nacional para la Cultura y las Artes, 2000.
- Calila* = *Calila e Dimna*. J. M. Cacho Blecua, & M. J. Lacarra (Eds.). Madrid: Castalia, 1984.
- Celestina* = *Fernando de Rojas, La Celestina*. D. S. Severin (Ed.), Madrid: Cátedra, 1993.
- Cid* = *Cantar de mio Cid. Texto, gramática y vocabulario*, Vol. 3: *Texto*, R. Menéndez Pidal (Ed.), Madrid: Espasa Calpe, 1944–1945.
- CN/Sí* = Leandro Fernández de Moratín, *La comedia nueva. El sí de las niñas*. J. Dowling, & R. Andioc (Eds.). Madrid: Castalia, 1968.
- Corbacho* = Alfonso Martínez de Toledo, *Arcipreste de Talavera o Corbacho*. M. Gerli (Ed.). Madrid: Cátedra, 1992 [1979].
- Davies, M. (2002–2016). *Corpus del Español: 100 million words, 1200s–1900s*. Retrieved from <<http://www.corpusdelespanol.org>>
- DLNE* = *Documentos lingüísticos de la Nueva España. Altiplano central*. C. Company Company (Ed.). Ciudad de México: Universidad Nacional Autónoma de México, 1994.
- Fazienda* = *La Fazienda de Ultramar*. D. Arbesú (Ed.). Retrieved from <<http://www.lafazienda-deultramar.com>>
- GE* = Alfonso X el Sabio, *General estoria, primera parte*. A. Solalinde (Ed.). Madrid: Centro de Estudios Históricos, 1930.
- GEI* = Alfonso X el Sabio, *General estoria. Primera parte*. L. Kasten, J. Nitti, & W. Jonxis-Henkemans (Eds.), *The electronic texts and concordances of the prose works of Alfonso X, El Sabio*. Madison: Hispanic Seminary of Medieval Studies, 1997.
- LT* = *Tri-linear edition of Lazarillo de Tormes of 1554 (Burgos, Alcalá de Henares, Amberes)*. J. V. Ricapito (Ed.). Madison: The Hispanic Seminary of Medieval Studies, 1987.

- Lucanor* = Don Juan Manuel, *El conde Lucanor o Libro de los enxiemplos del conde Lucanor et de Patronio*. J. M. Blecua (Ed.). Madrid: Castalia, 1971.
- Madrid* = *El habla de la ciudad de Madrid: materiales para su estudio*. M. Esgueva, & M. Cantarero (Eds.). Madrid: Consejo Superior de Investigaciones Científicas, 1981.
- México* = *El habla de la ciudad de México. Materiales para su estudio*. J. M. Lope Blanch (Ed.). Ciudad de México: Universidad Nacional Autónoma de México, 1971.
- Quijote* = Miguel de Cervantes, *Don Quijote de la Mancha*. In *Obras completas*, F. Sevilla Arroyo, & A. Rey Hazas (Eds.). Madrid: Alianza Editorial-Centro de Estudios Cervantinos, 1996.
- Regenta* = Leopoldo Alas "Clarín", *La Regenta*. G. Sobejano (Ed.). Madrid: Castalia, 1981–1982.
- Zifar* = *El libro del cavallero Zifar*. Ch. Ph. Wagner (Ed.). Ann Arbor: University of Michigan Press, 1929.

Recurrent processes in the evolution of concessive subordinators in Spanish and Catalan

Manuel Pérez Saldanya & José Ignacio Hualde

Universitat de València / University of Illinois at Urbana-Champaign

Concessive conjunctions show a tendency to be often recycled and replaced along certain specific grammaticalisation patterns. Here, after examining those general properties of concessive conjunctions that explain their historical instability, we focus on the history of two such elements in Spanish and Catalan. First, we trace the semantic development of Spanish *aunque*/Catalan *encara que* from conditional concessive ('even if') to purely concessive ('although') and even, in some contexts, adversative ('but'). Secondly, we consider the trajectory of Spanish *puesto que*, which is first attested with conditional concessive value ('provided that', 'even if') and only later became a causal conjunction ('since'), apparently at odds with the standard pattern of development that we see in other cases.

Keywords: concessive; conditional; causal; subordination; grammaticalisation; Spanish; Catalan

1. Introduction

The development of concessive conjunctions is a classical topic in historical syntax. One reason that explains the interest of historical linguists in this topic is the fact that concessive conjunctions tend to undergo cyclic recreation. Some categories tend to be diachronically stable and others, instead, are subject to recurrent change. For instance, among subordinating elements, the stability of the conditional particle *si* 'if' in the history of the Romance languages is in contrast with the great instability shown by the concessive conjunctions, which have undergone frequent replacement (see Rodríguez Molina & Enrique-Arias, in this volume, for more

information on the Old Spanish use of *si* as a Q-particle). In (1), we schematise the pattern of replacement in basic concessive conjunctions from Latin to present-day Spanish and Catalan:

- (1) Basic concessive conjunctions from Latin to Sp. and Cat.:
- | | | |
|--------------------------------------|---|------------------------|
| Lat. QUAMVIS | → | LLat. QUAMLIBET |
| → OSp. <i>maguer</i> (<i>que</i>) | → | Sp. <i>aunque</i> |
| → OCat. <i>jatsia</i> (<i>que</i>) | → | Cat. <i>encara que</i> |
| 'Although' | | |

The Classical Latin concessive conjunction QUAMVIS, which competed with QUAMQUAM and several other forms (Martín Puente, 2002), was replaced in Late Latin by QUAMLIBET, which, in turn, was replaced by *maguer* (*que*) in Old Spanish (Montero Cartelle, 1992c) and by *jatsia* (*que*) in Old Catalan (Rofes, 2007, 2010). At the end of the Middle Ages we find a new replacement of this concessive conjunction as Spanish develops *aunque* as its main concessive conjunction and Catalan *encara que*, although other concessive expressions reached very high frequencies for a while, such as *comoquier que* in Spanish (Montero Cartelle, 1992a, 1992b) or *si bé* and (*per*) *bé que* in Catalan (Rofes, 2012a).

A second property that may also explain the interest that concessive conjunctions have generated is the fact that they tend to arise from rather transparent lexical or grammatical sources (König, 1988). Both Latin conjunctions QUAMVIS and QUAMLIBET derive from constructions with verbs of will and desire: VIS 'you want', and LIBET 'it pleases'. The Old Spanish conjunction *maguer* arose from an exclamative expression, borrowed from Greek *μακάρι* 'hopefully', also with desiderative value. The remaining forms developed from nonfactual constructions with an aspectual adverb ('already'; 'still', 'even'):

- (2) a. QUAMVIS: 'as you want' > 'although'
 b. QUAMLIBET: 'as it pleases one' > 'although'
 c. OSp. *maguer* 'hopefully' (< Gr. *makári*) > 'although'
 OCat. *jatsia* 'be it already' (< *IAM SE SIAT) > 'although, albeit'
 d. Sp. *aunque* (< *aún que* 'still + that') > 'even that' > 'although'
 Cat. *encara que* 'still + that' > 'even that' > 'although'

Finally, and related to the preceding point, we may add that the same lexical sources tend to recur in different languages, without this being necessarily due to language contact. Given this diachronic and synchronic richness, it is not surprising that concessive constructions have been the object of considerable attention within historical and typological linguistic studies (Harris, 1988; Haspelmath & König, 1998; König, 1985a, 1985b; Salvador, 2010).

In this paper we attempt to explain the inherent instability of the concessive conjunctions, focusing on some recurrent patterns in the origin and evolution of concessive conjunctions in Spanish and Catalan. First, we analyse the main properties of concessive constructions, and the lexical and grammatical sources of concessive conjunctions in these two languages (Section 2). Secondly, we will consider the development of Sp. *aunque*/Cat. *encara que*, which follow a common diachronic path whereby conditional concessives become pure concessives and may even develop an adversative value in some contexts (Section 3). Finally, after reviewing some cases in Spanish and Catalan illustrating the common evolution from causal to concessive meaning (Section 4.1), we will discuss the unexpected change in the opposite direction that we observe in the history of Sp. *puesto que* (Section 4.2).

2. Main properties of concessive constructions

2.1 Concessives as semantically complex constructions

We may ask now why it is that concessive conjunctions tend to be cyclically replaced or why they do so at a higher rate than other subordinators. A first answer may be related to the semantic and pragmatic properties that concessive constructions display. Semantically, these are constructions that have a complex or “derived character” (König, 1988). Pragmatically, the expression of concessivity implies, as we shall see, a “pragmatic cost”, which the speaker may try to minimize by making use of new and more expressive resources, having a greater impact on the listener.

Concessive clauses have a complex and derived character because they express a causal or conditional condition, albeit through a heterogeneous relation that negates the causal or conditional link. Compare the structure of simple causal and conditional relations in (3a) and (3b), respectively, with the meaning structure of the concessive examples in (3c) and (3d):

- (3) a. *Causal*: Since he asked me [*cause*], I told him everything [*effect*]
 b. *Conditional*: If he asks me [*condition*], I will tell him everything [*conditioned*]
 c. *Factual concessive*: Even though he asked me [*inefficient cause*], I did not tell him anything
 d. *Nonfactual concessive*: Even if he asks me [*inefficient condition*], I will not tell him anything

Whereas in causal and conditional constructions main and subordinate clauses have the same argumentative orientation, in concessive constructions there are

opposite orientations in the two clauses (see Anscombe & Ducrot, 1983 on argumentative orientation). Concessive constructions presuppose the assumption of a causal or conditional relation, which is then negated. Concessives thus have a counterexpectative and counterargumentative character, and this has a pragmatic cost that must be minimised.

There are two main types of concessive constructions: those that presuppose a causal relation (from actual cause to actual effect), as in (4), and those that presuppose a conditional relation (from possible or counterfactual condition to possible or counterfactual effect), as in (5). The former are known as pure or proper concessives and the latter as conditional concessives (see König, 1985a, 1985b, 1988; Quer, 1988, pp. 235–260; Rodríguez Rosique, 2008, pp. 213–267).

- (4) Pure concessives (= 3c)
 PRESUPPOSED: Usually, given p, q
 ASSERTED: But in this case, p & not q
- (5) Conditional concessives (= 3d)
 PRESUPPOSED: Usually, if p, then q
 ASSERTED: But in this case, if p, not q

In Table 1, adapted from Rofes (2012a, Section 2.2.2), the two types of concessives are compared with causals and conditionals.

Table 1. Causal, conditional and concessive constructions

Argumentative orientation	Truth value of subordinate clause	
	Factual	Nonfactual (or counterfactual)
Harmonic	Causal	Conditional
Disharmonic	(Pure) concessive	Concessive conditional

In English, the distinction between pure concessives and conditional concessives is expressed by different subordinators: *although*, *though*, *in spite of* in pure concessives and *even if* in conditional concessives. In the languages we are concerned with in this paper, Spanish and Catalan, on the other hand, the choice of conjunction does not establish such a clear contrast, since some conjunctions may have both values. Instead, the distinction is made in part through a choice of mood in the subordinate: indicative or subjunctive.

With ambivalent conjunctions, such as Spanish *aunque* ‘although/even if’ and its Catalan counterpart *encara que*, the indicative mood has a factual character, giving rise to a pure concessive interpretation, as in (6a), whereas the subjunctive’s

nonfactual value produces a conditional interpretation, as in (6b). Nevertheless, it should be remembered that the subjunctive mood is not incompatible with a factual value. This is so in contexts where the subordinate clause is presented as having little informative relevance as effective cause, as in (6c).¹

- (6) a. *Aunque lo sabe, no nos dirá nada* (Sp.)
Encara que ho sap, no ens dirà res (Cat.)
 ‘Even though s/he knows it, s/he won’t tell us anything’
- b. *No sé si lo sabe, pero aunque lo sepa, no nos dirá nada* (Sp.)
No sé si ho sap, però encara que ho sàpiga, no ens dirà res (Cat.)
 ‘I don’t know if s/he knows it, but even if s/he knows it, s/he won’t tell us anything’
- c. *Claro que lo sabe, pero aunque lo sepa, no nos dirá nada* (Sp.)
És clar que ho sap, però encara que ho sàpiga, no ens dirà res (Cat.)
 ‘Of course s/he knows it, but even though s/he knows it, s/he won’t tell us anything’

2.2 Other properties of concessive constructions

As already mentioned, concessive clauses negate a causal or conditional relation. From a diachronic point of view, in a first stage, it is common for causal or conditional constructions to acquire a contextual concessive value when the main clause contains negation or a contrastive element. In (7)–(9) the use of several elements that activate a concessive interpretation is illustrated with examples from Old Spanish (7)–(9a) and Old Catalan (7)–(9b). Negation is the most important concessive activator, as in (7), but the contrast required for a concessive interpretation to arise can also be established by means of other elements such as comparatives of inequality, as in (8), or antithetical terms, as in (9). These are all different elements that explicitly indicate that main and subordinate clause have opposite argumentative orientations.

- (7) a. *maguer los están llamando, ninguno non responde*
 ‘Even though they are calling them, nobody responds’
 (OSp.; *Cid*: 2305, ca. 1140)

1. This use of the subjunctive has been labelled factual, thematic and polemic. The term *factual* is justified by the fact that the subordinate clause refers to an event that has taken place. The term *thematic* indicates that the clause includes given information. Finally, the label *polemic* refers to the fact that the validity of the information contained in the subordinate clause is negated. See Vallejo (1922, pp. 49–50), Rivarola (1976, pp. 12–13), Pérez Saldanya (1999, section 50.2.3.3) and Montero (2002).

- b. *ja fos so que Senta Maria fos molt bela, negú no se'n podia enamorar ni la poc cobeseyar desaordenadament*
 'Even though St. Mary was very beautiful, nobody could fall in love with her or desire her improperly' (OCat.; *Vides*: 252, XIII)
- (8) a. *Maguer que era ante por precioso contado,/desende adelante fo mucho máspreciado*
 'Although it was already taken as precious before, from that moment on it was even more valued' (OSp.; *Santo Domingo*: 192, ca. 1236)
- b. *si ben se fo menor de totz los apòstols, emperò més trebaylà que totz*
 'Even though he was the smallest of all the apostles, he worked harder than all the others' (OCat.; *Diàlegs*: 17r)
- (9) a. *Maguer plogo al rey, mucho pesó a Garcí Ordóñez*
 'Even though it pleased the king, it caused Garcí Ordóñez much sorrow' (OSp.; *Cid*: 1345, ca. 1140)
- b. *E ja fos en temps de gener, que fa gran fret, contornam-nos la nuyt més de ·C· vegades e-l lit, de la I^a part e de l'altra, e suàvem també con si fóssem en ·I· bany*
 'And even though it was the month of January, when it is very cold, we turned around in bed more than one hundred times in the night, from one side to the other, and we were also sweating, as if we were in a bath' (OCat.; *Fets*: 106r, 1343)

Concessive constructions share their contrastive value with adversative clauses. In both types of sentences there is a contrast, a relation between thesis and antithesis. The difference is found in their focus. In concessive constructions, the focus of the contrast is on an inefficient cause or condition, expressed by means of the subordinate clause, which is generally preposed and functions as a theme (10a). Adversative constructions, on the other hand, have a coordinated structure and highlight the unexpected effect, which has a rhematic value (10b). Nevertheless, if a concessive clause is postposed to the main clause and is separated from it by a pause or an intonational break, it may be very close in meaning to an adversative clause (10c).

- (10) a. Even though he knew it [*theme*], he didn't tell her
 b. He knew it, but he didn't tell her [*rheme*]
 c. He didn't tell her, even though he knew it [*rheme, concessive with adversative value*]

2.3 Lexical and grammatical sources of concessives

Most diachronic studies on concessive constructions have focused on their lexical and grammatical sources, pointing out their general, universal character (Harris, 1988; Haspelmath & König, 1998; König, 1985a, 1985b; Salvador, 2010). The following sources of concessives are well established: (a) causal and conditional

constructions, (b) constructions with a universal quantifier or an adverb of inclusion that refers to the highest point within a scale, and (c) constructions expressing intention, desire or will that an event should happen or fear that it may happen. These sources are illustrated for Spanish and Catalan in Table 2.

Table 2. Concessive conjunctions classified by their source

Type of construction	Spanish	Catalan
Causal construction	†(<i>em</i>) <i>pero que</i> <i>por más que</i>	<i>per més que</i> <i>per bé que</i>
Conditional construction	<i>si bien</i> <i>siquier(a)</i>	<i>si bé</i>
Construction with a universal quantifier 'all'	<i>con todo (y) que</i>	(<i>amb</i>) <i>tot (i) que</i>
Construction with a focal element of inclusion 'even'	<i>aunque</i>	<i>encara que</i>
Construction of desire or intention	† <i>maguer (que)</i> † <i>comoquier(a) que</i> <i>siquier(a)</i>	† <i>jatsia (que)</i> <i>baldament</i>
Construction with adverb of assertion	<i>bien que</i> <i>si bien</i>	(<i>per</i>) <i>bé que</i> <i>si bé</i>
Construction with element of disapproval	<i>mal que</i> <i>a pesar de que</i> <i>pese a que</i>	<i>malgrat que</i>

It is clear that causal and conditional constructions may readily develop into concessives, since the only difference is found in their argumentative orientation. Regarding other sources, the relation is more indirect. The explanation for why these specific sources give rise to concessive conjunctions is to be found in the diverse strategies that a speaker may use in order to minimise the pragmatic cost of negating a causal or conditional relation that is presupposed as normally holding. This may be assumed to be costly for the listener because of the complex processing of inferences that is required.

The speaker may refer to the set of all possible situations (e.g., Cat. *tot i que*, lit. 'all and (that)'), or to the optimal situation within a scale of possible situations (e.g., Sp. *aunque*, lit. 'even (that)'), in order to make clear that the causal/conditional relation does not hold in a specific case. The speaker may also empathise with the listener by expressing a desire or acceptance of an event taking place (e.g., Sp. *siquiera*, lit. 'if one may want') or the fear or regret that it may take place (e.g., Sp. *a pesar de que*, 'in spite of', lit. 'with sorrow, regret that').

As indicated in the introduction, in the remainder of this paper we will analyse two sources of concessives in more detail. First, we will consider the development of Sp. *aunque*/Cat. *encara que*. Secondly, after reviewing some cases in

Spanish and Catalan illustrating the common evolution from causal to concessive meaning, we will discuss the unexpected change in the opposite direction that we observe in the history of Sp. *puesto que*.

3. Spanish *aunque*/Catalan *encara que*: From conditional concessive to adversative

The Spanish concessive conjunction *aunque* ‘although’, which in some contexts can be used with a purely adversative value, has a transparent source in the adverb *aun* ‘even’ (which in turn derives from the time adverb *aún* ‘still’). Cat. *encara que* is its direct equivalent (cf. *encara* ‘still’). In this section we trace the development of this pair of conjunctions in some detail.

3.1 Scalar adverbs in the development of concessive conjunctions

Adverbs of inclusion such as Sp. *incluso* and *aun* ‘even’ are focalising elements that may modify any constituent in the sentence, including, among others, prepositional phrases, as in (11a), noun phrases, as in (11b), and temporal subordinate clauses, as in (11c):

- (11) a. *Pedro me habló incluso/aun de sus problemas más personales*
 ‘Pedro spoke to me even about his most personal problems’
 b. *Asistirán a la boda incluso/aun los primos de México*
 ‘Even our cousins from Mexico will come to the wedding’
 c. *Van de excursión a la montaña incluso/aun cuando llueve*
 ‘They go hiking in the mountains even when it rains’

These focal adverbs are scalar. They presuppose the existence of a set of entities, properties or events that includes the one designated by the focalised constituent and indicate that the said entity, property or event is located at one end of a graded scale (see Gast & van der Auwera, 2011 for a more detailed analysis). In (11a), for instance, the set that is evoked includes all topics mentioned by Pedro. Within this set, the topic of Pedro’s personal problems is the most unexpected one and thus the one that carries the greatest informative load. Because of their scalar nature, these focal adverbs can be used in concessive constructions when the constituent focalised by the adverb is an adjunct with conditional value, be it a finite clause, as in (12a), a non-finite clause, as in (12b), or a prepositional phrase, as in (12c).

- (12) a. *Incluso si lo propone él, nos lo denegarán*
 ‘Even if he proposes it, it will be denied to us’

- b. *Incluso proponiéndolo él, nos lo denegarán*
‘Even with him proposing it, it will be denied to us’
- c. *Incluso con él, nos lo denegarán*
‘Even with him, it will be denied to us’

In these scalar concessive constructions, the focalised constituent refers to an event, entity or property that is introduced as optimal in order to obtain a given result, which, nevertheless, turns out to be ineffective (see Rodríguez Rosique, 2008, pp. 269–341 for a detailed analysis of these constructions in Spanish). The fact that the optimal event is ineffective implies the inefficacy of all other possible events, which gives more expressive force to the argument.

3.2 From *aun + que* ‘even + that’ to *aunque* ‘although’

Spanish *aunque* ‘although’ is a grammaticalised concessive conjunction with a scalar origin, *aun que* ‘even (that)’, ultimately from a time adverb ‘still’. Several hypotheses have been put forward on the origin of *aunque*. The most common view is that it arose in clauses with concessive *que* ‘that’ reinforced by the adverb *aun* ‘even’ (Bartol, 1986, pp. 201–203; Cortés, 1992, pp. 187–191; Rivarola, 1976, pp. 201–203). A problem with this hypothesis, however, is that the concessive value of *que* is purely contextual and only sporadic. It seems more sensible to see the origin of *aunque* in focal uses of *aun* ‘even’ where this adverb modifies a subordinate clause headed by *que* (see Elvira, 2005; Rodríguez Rosique, 2008, pp. 344–345; for Catalan see Pérez Saldanya & Salvador, 1995).

The appearance of the adverb *aun* in concessive constructions is consistent with a focal usage that is already found with the Latin time adverb ADHUC ‘until now’, from which Sp. *aún* ‘still’, *aun* ‘even’ developed (notice the difference in spelling and pronunciation between *aun* ‘even’ and *aún* ‘still’ in present-day Spanish). Alongside *aunque*, we find *aun si* ‘even if’, *aun cuando* ‘even when’, etc., as in (13). The fact that *aunque* is written as a single word reflects a greater degree of grammaticalisation of this expression.

- (13) a. *E allí fallaron un campo tan llano e tan bueno e tan grand que estando en medio d’él que non podién devisar los cabos, e semejóles que allí cabrién todos, e aun si más fuessen*
‘And there they found a field so flat, and so good, and so big, that, standing in the middle, they could not see the borders; and it seemed to them that they would all fit there, and even if there were more of them’
(*General estoria*: 1.74, 1270–1280)
- b. *De mí non puedes nula cosa levar, aun porque quisiesse non terría qué dar*
‘You cannot take anything from me; because even if I wanted, I would not have anything to give’
(*Santo Domingo*: 95, ca. 1236)

- c. *que quiere él que ayan su virtud e su poder e su onra, e però esto se cumple aun cuando se faze el esperamiento con el ordenamiento de las estrellas que pertenecen a ello, e es estonces el tiempo*
 ‘That he wants them to have their virtue, and their power, and their honour; but this is accomplished *even when* one waits for the arrangement of the stars that corresponds to it, and then it is time’
 (General estoria: 1.76, 1270–1280)
- d. *Porque, aun desque vos murierdes, siempre viva la fama de los vuestros fechos*
 ‘So that, *even after* you die, the fame of your deeds may live forever’
 (Lucanor: 117, 1335)

It thus seems reasonable to postulate that the conjunction *aunque* with concessive value arose in contexts where a focal adverb *aun* ‘even’ was followed by a subordinate clause headed by *que* ‘that’, as in (14), where *aunque* appears in the same sentence as the older concessive conjunction *maguer*. In this example, both *maguer que* and *aunque* have a conditional concessive value, but the latter introduces a more optimal situation to obtain the result of “understanding”, and its scalar and conditional concessive value is, therefore, clearer and stronger.

- (14) *confondámosles el language que an agora todos uno, e mezcámosgele de guisa que maguer que se oyan que se non entiendan aunque estén muy decerca unos d’otros*
 ‘Let us confound their language, that they now all have one, and let us mix it in such a way that, *even if* they hear each other, they are not able to understand one another, *even if* they are very close to each other’
 (General estoria: 1.76, 1270–1280)

3.3 Final stages in the grammaticalisation of *aunque*

Aunque is first attested in the second half of the 13th century. Initially, it had a conditional value, as shown in example (14). The subsequent development of Sp. *aunque* follows a well-known path of grammaticalisation, according to which the contrast introduced by the scalar particle *aun* ‘even’ enables its drift towards exclusively concessive value (Rodríguez Rosique, 2008, pp. 273, 343–350). When we consider both the mood of the clause introduced by *aunque* and the position of this clause before or after the main clause, we find significant changes in relative frequency from the 13th century, when this concessive conjunction is first attested, to the 17th century, when it becomes the most important concessive conjunction in Spanish. Table 3 shows the results of the analysis of several representative texts from this period. In this table, besides separating by position, we classify *aunque*-clauses as containing a verb in the indicative mood (“ind.”), in the subjunctive with factual meaning (“subj. f”), in the subjunctive with nonfactual meaning (“subj. nf”), and not containing a verb (“w/o V”).

Table 3. Subordinate clauses with *aunque*

Text and year	Preceding the mainclause				Following the mainclause			
	Ind.	Subj. f	Subj. nf	W/o V	Ind.	Subj. f	Subj. nf	W/o V
<i>GEI</i> 1270–1280			60% (6/10)		10% (1/10)		30% (3/10)	
<i>Lucanor</i> 1335			70.8% (17/24)				29.1% (7/24)	
<i>Pedro I</i> 1370	12.5% (1/8)		25% (2/8)		25% (2/8)		37.5% (3/8)	
<i>Generaciones</i> 1450–1455	23.5% (4/17)		23.5% (4/17)		47% (8/17)		5.8% (1/17)	
<i>Enrique IV</i> 1480	27% (30/111)	9.9% (11/111)	7.2% (8/111)	10.8% (12/111)	22.5% (25/111)	1.8% (2/111)	8.1% (9/111)	12.6% (14/111)
<i>Celestina</i> 1499	16.6% (18/108)	8.3% (9/108)	27.7% (30/108)	3.7% (4/108)	21.3% (23/108)	5.5% (6/108)	12% (13/108)	4.6% (5/108)
<i>Diálogo</i> 1535	36.4% (28/77)		15.6% (12/77)		31.1% (24/77)	3.9% (3/77)	11.7% (9/77)	1.3% (1/77)
<i>Coloquios</i> 1547–1551	13.7% (16/117)	6% (7/117)	26.5% (31/117)	2.6% (3/117)	30.8% (36/117)		17% (20/117)	3.4% (4/117)
<i>Bernal</i> ca. 1568	31.7% (40/126)	1.6% (2/126)	15% (19/126)	3.2% (4/126)	25.4% (32/126)		15% (19/126)	8% (10/126)
<i>Quijote I</i> 1605	27.4% (74/270)	3.7% (10/270)	12.2% (33/270)	8.5% (23/270)	21.8% (59/270)		21.5% (58/270)	4.8% (13/270)
<i>Hombre</i> 1686	14.4% (16/111)	13.5% (15/111)	37.8% (42/111)	4.5% (5/111)	7.2% (8/111)		15.3% (17/111)	7.2% (8/111)

In the first stage, *aunque* has a conditional concessive value and, consequently, the clause that it heads has a nonfactual interpretation, with the verb almost always in the subjunctive mood (as in (14) above). In fact, with very few exceptions, concessive clauses with *aunque* appear exclusively in the subjunctive mood until the first half of the 14th century. The subjunctive is still more frequent than the indicative in clauses with *aunque* during the second half of this century. Progressively, however, this conjunction loses its scalar value, becomes more frequent, and then the use in indicative clauses (15a) or even without a verb (15b) becomes normal.

- (15) a. *En mi vida me acuerdo aver tan gran temor, ni verme en tal afrenta, aunque he andado por casas ajenas harto tiempo y en lugares de harto trabajo*

‘I don’t remember feeling such great fear in my life or seeing myself in such hardship, *even though* I have been in other people’s homes and in places of much work for a long time’ (Celestina: 264, 1499)

- b. *Pero con todo eso çerca del rey don Alfonso avia muy nobles onbres e muy estrennos, e aunque en numero pocos en virtud eran grandes*
 ‘But nevertheless there were very noble and distinguished men in King Alfonso’s entourage and, *even though* [they were] few in number, they were great in virtue’
 (Enrique IV: 209, ca. 1481–1482)

Another indication of degree of grammaticalisation is freedom of position with respect to the main clause. As is well known, one of the main differences between adversatives and concessives is freedom of position: *Juan trabaja mucho, pero es pobre* ‘Juan works hard, but is poor’ ~ **Pero es pobre, Juan trabaja mucho* ‘*But he is poor, Juan works hard’, *Aunque trabaja mucho, Juan es pobre* ‘Even though he works hard, Juan is poor’ ~ *Juan es pobre, aunque trabaja mucho* ‘Juan is poor, even though he works hard’. Although, most commonly, concessive clauses tend to appear before the main clause, a great degree of grammaticalisation may allow for their collocation after the main clause.

At an even more advanced stage of the grammaticalisation process, concessive conjunctions may develop a restrictive adversative value. This value may arise in indicative clauses following the main clause and separated from it by a pause or an intonational break. This adversative value is common from the second half of the 15th century on. Examples are given in (16):²

- (16) a. *Yo veo bien mi duelo, aunque me lo callo*
 ‘I feel my pain plainly, *even if/but* I do not mention it’
 (Corbacho: 126, 1438)
- b. *agora pues come pecador, que, si a dios plazze, presto nos veremos sin necesidad, aunque te digo que despues que en esta casa entre, nunca bien me ha ydo*
 ‘Eat now, sinner, because, if it pleases God, we will soon see ourselves free of need, *although/but* I confess that, from the moment I entered this house, things have never gone well for me’
 (Lazarillo: 48, 1554)
- c. *Bien dices, Pedro – dixo uno de ellos –, aunque no será menester usar de esa diligencia, que yo me quedaré por todos*
 ‘You are right, Pedro – said one of them – *although/but* it will not be necessary to do it that way, because I will stay here for all’
 (Quijote I: 129, 1605)

2. As Rivas (1989) argues, the use of *aunque* in such contexts can be interpreted as a contextual neutralisation of the distinction between concessive and adversative value.

3.4 Catalan *encara que*

The Catalan conjunction *encara que* has exactly the same development as we have just described for Spanish *aunque*, but the process is a little slower and *encara que* does not replace older *jatsia* until the end of the Middle Ages (Rofes, 2007, 2012a, Section 3.2.2). In (17) we offer an early Catalan example of the concessive conjunction *encara que*.

- (17) *deym que les pres per sa pròpria actoritat, e encara que les presés permanament de son seynor ne de justícia, ço que no féu gens, per ço no romanrie que el dit Bernat no fos tengut de les dites peynores a redre o de pendre en son comte del deute per él demanat, pus que per rahó de deute les prenie*

‘We say that he took them by his own authority; and *even if* he had taken them by command of his lord or by justice, which he didn’t do at all, it would not follow that the said Bernat did not have the right to return the said pledges or to receive them in consideration of the debt he was requesting, since he took them because they were owed to him’

(*Cocentaina*: 33, 1269)

The historical development of *encara que*, from appearing exclusively in subjunctive clauses to its gradual extension to indicative and other contexts, is traced in Table 4.

Table 4. Subordinate clauses with *encara que*

Text and year	Preposed concessive				Postposed concessive			
	Ind.	Subj. f	Subj. nf	W/o V	Ind.	Subj. f	Subj. nf	W/o V
<i>Cocentaina</i> 1269–1290			85.7% (6/7)				14.2% (1/7)	
<i>Furs</i> 1330			66.6% (4/6)				33.3% (2/6)	
<i>Corbatxo</i> f. xiv		20% (1/5)	60% (3/5)				20% (1/5)	
<i>Alcavota</i> 1410			100% (2/2)					
<i>Tirant</i> 1490	1.7% (2/113)	27.4% (31/113)	31.8% (36/113)		2.6% (3/113)	4.4% (5/113)	31.8% (36/113)	
<i>Vida</i> 1515	4% (2/49)	26.5% (13/49)	38.7% (19/49)		10.2% (5/49)	6.1% (3/49)	10.2% (5/49)	4% (2/49)
<i>Generalitat</i> 1509–1539		11.1% (1/9)	44.4% (4/9)		44.9% (4/9)			

(Continued)

Table 4. (Continued)

Text and year	Preposed concessive				Postposed concessive			
	Ind.	Subj. f	Subj. nf	W/o V	Ind.	Subj. f	Subj. nf	W/o V
<i>Col·loquis</i> 1557	29.7% (11/37)	16.2% (6/37)	13.5% (5/37)	10.8% (4/37)	10.8% (4/37)	2.7% (1/37)	16.2% (6/37)	
<i>Memòries</i> 1572–1602	30.7% (4/13)		30.7% (4/13)		53.8% (7/13)		7.6% (1/13)	
<i>Dietari</i> 1589–1629	10.6% (8/75)		4% (3/75)	1.3% (1/75)	64% (48/75)		17.3% (13/75)	2.6% (2/75)

At a first stage, *encara que* appears in subjunctive clauses with nonfactual value and is found only in legal texts such as charters and court of law proceedings. In the second half of the 15th century, we start finding this concessive conjunction in indicative clauses, as in the example in (18a), although the subjunctive use is still dominant in works of this period such as *Tirant lo Blanch* and *Curial* (Rofes, 2012a, Section 3.2.2). From the 16th century on, when *encara que* becomes the most frequent concessive conjunction, we see an increase in its use in indicative clauses and following the main clause, as well as in constructions without an explicit verb, as in (18b). From that time on, it is also possible to find examples with adversative value, as in (18c) and (18d).

- (18) a. *E encara que ella tenia molt bella persona e ben disposta, emperò les calces vermelles e lo capell al cap la desfavoria tant que paria que fos un diable*
‘And, although she had a very beautiful and well-proportioned body, nevertheless, the red stockings and the hat on her head were so unflattering that she looked like a devil’ (Tirant: 897, 1490)
- b. *Desijós prèns esforç y, refocil·lat algun poquet, levà·s de terra e respongué a Amor·de·Déu, encara que ab molts sanglots y veu regullosa y temorejant*
‘Willingly he made an effort and taking courage he got up from the ground and replied to Love-of-God, although with many tears and with a fearful voice’ (Vida: 41r, 1515)
- c. *E, per causa que lo reverent archbisbe de Tarragona prengué en aquest dia lo pal·li en lo monestir de Pedralbes, manchà molta gent de la convidada en la dita festivitiat, encara que en les segones vespres hi hagué tanta multitut de poble y, especialment, dones que no y havia loch a hon poder estar ab pler*
‘And because the reverend archbishop of Tarragona on that day received the pallium in the monastery of Pedralba, many people among those invited to the said festivity were absent, although/but on the second eve there were so many people there, especially women, that there was no space where one might be at ease’ (Generalitat: 410b, 1531)

- d. *Jo tindrè diligència en sollicitar la espedició, enca[ra] que no és menester, que don Juan, mon senyor, ne té molt-ància*
 ‘I will make sure to ask for it to be produced, although/but it is not necessary, because Don Juan, my lord, is very eager for it’
 (*Epistolaris*: 107, 1534)

4. From causal to concessive (and vice versa?)

Crosslinguistically, causal constructions are a common source of concessive conjunctions. In Spanish a conditional expression with this origin is *por (más) que* ‘no matter how (much)’, whose development we trace here. The pattern of development from causal to concessive is well understood and is expected to be unidirectional. Nevertheless, we also find cases like that of Sp. *puesto que* ‘since’, where we seem to have the opposite development. We argue that, rather than dealing with a process of grammaticalisation, what we have here is a reinterpretation made possible by the different meanings that absolute clauses with the participle *puesto* ‘put’ may acquire in context.

4.1 Spanish *por más que*/Catalan *per més que*: From causal to concessive value

The semantic link between causal and concessive clauses is clear in examples such as the well-known verses by Sem Tob de Carrión in (19).

- (19) *Por nasçer en espino/non val la rosa, çierto,/menos, nin el buen vino/por salir del sarmiento;/non val el açor menos/por nasçer de mal nido,/nin los exenplos buenos/por los dezir judío*
 ‘For having been born amidst thorns, the rose is not worth less and neither is good wine for coming from a vine shoot; a goshawk does not have less value for having been born in a bad nest, nor do good fables for being told by a Jew’
 (Sem Tob, *Proverbios*: 136–137, ca. 1355)

In (19), the causal clauses with *por* ‘for’ acquire a concessive meaning in the context of these verses because the negation breaks the causal link between main and subordinate clause, giving rise to a non-effective cause interpretation (see Section 2.2). This contextual value may become fixed by a process of routinisation, so that the presence of the negative element is no longer necessary. In both Spanish and Catalan we find more-or-less grammaticalised concessive constructions with a causative source.

In Spanish the contextual concessive value of ponderative structures of the type *por mucho/más/bien... que* ‘no matter how much/well, etc.’ has been grammaticalised. An Old Spanish example is given in (20a). We also find the same grammaticalisation of concessivity in Catalan *per més/molt/bé... que*, as in (20b).

- (20) a. *et plógol más porque açertó en l' su consejo que sil oviera acaesçido otra pro o otra onra por grande que fuesse*
 'And he was more pleased because he was right in his advice than if he had received any other benefit or honour, no matter how great'
 (Lucanor: 155, 1335)
- b. *Tant fo gran lo meravelós goyg que nostra dona santa Maria ach, que tot hom, per molt que perda e per moltes tribulacions que age e per molta passió que sostenga, pot ésser consolat e alegrat e beneficiat en lo gog de nostra dona*
 'The marvelous joy that Our Lady St Mary had was so great that everyone, no matter how much he loses and no matter how many sorrows he has and no matter how much he suffers, can be consoled and rejoiced and take delight in Our Lady's joy'
 (Llull, *Doctrina*: 116, 14th c.)

In these constructions the constituent with quantitative value following the causal preposition is not in the position that we would expect from its function. Notice that in (20a), *grande* 'great' is a complement of *fuese* 'it were'. Similarly, in (20b) *molt* 'much' and *moltes tribulacions* 'many tribulations' are complements of *perda* 'he may lose' and *age* 'he may have', respectively. These constituents, however, are not found in postverbal position, but are preposed to the conjunction *que* 'that'. This is the same pattern that we find in emphatic constructions like those in (21), where *que* is optional.³

- (21) a. ¡*Qué grande (que) es!*
 'It is so big!'
- b. ¡*Menudo susto (que) nos diste!*
 'You gave us such a scare!'

As already pointed out by Vallejo (1922, pp. 40–48) and reiterated by more recent scholarship (Algeo, 1981; Elvira, 2003), the origin of the concessive structure illustrated in (20) is to be found in causal phrases where the complement of the preposition *por* 'for' is a bare noun phrase. Let us consider the Old Spanish example in (22):

3. Here we follow the analysis in Brucart (1993) and RAE/ASALE (2009, section 42.15), according to which these are constructions with anteposition and a conjunction *que* that signals emphasis. In concessive constructions the complementizer *que* also signals that the subordinate sentence is a complement of the preposition *por*.

- (22) *e por esta razón fizieron los reis aquesto, ca non por sus voluntades nin por sabor que oviessen de mudar e renovar fueros*
 ‘And for this reason did the kings do this; not because it was their will or because it was their pleasure to change and renew the laws’ (*General Estoria*: 2.600, 1270–1280)

In the example in (22) *por sabor que oviessen* ‘for (the) pleasure they had’ can be read as equivalent to *por mucho sabor que oviessen* ‘for much pleasure they had’, which, in the context of the negation, admits a concessive interpretation. The hypothesis is that the contextual value of scalar concessive construction became fixed, causing the reanalysis of the construction along the same pattern as the emphatic construction in (21). The proposed reanalysis is schematised in (23). In this example the complement of the preposition *por* ‘for’ is syntactically reanalysed from noun phrase to subordinate clause. The relative pronoun *que* ‘which, that’ is reinterpreted as a non-relative complementiser and the noun *sabor* ‘pleasure’ is understood as occupying a focus position preceding the complementiser.

- (23) [*por* [_{NP} *sabor*_i [_S*que*_i *oviesen* \emptyset _i]]
 → [*por* [_S [_{FOC} *sabor*_i] [*que* *oviesen* \emptyset _i]]

The existence of this reanalysis is already apparent in texts from the first half of the 14th century in both Spanish and Catalan. This is shown by the presence of the construction in non-negative contexts where *por/per* has complements that are unlikely for a causal interpretation, as in the two examples in (20) above. Around the same time we also find the first examples with a verb in the indicative mood and with factual value, as in (24), according to the tendency for conditional concessive constructions to acquire a purely concessive value.⁴

- (24) a. *Et por mucho que los llamó, non respondió ninguno dellos, que eran ydos todos*
 ‘And no matter how much he called them, none of them answered, because they had all left’
 (*Lucanor*: 269, 1335)

4. This shift in mood is parallel to that noticed for the constructions analysed in Section 3. However, in these constructions with a causal origin, the use of the indicative mood has remained in the minority, because the construction has retained its scalar value, which is associated with conditional concessives. We find only sporadic cases of *por más que* (and Cat. *per més que*) without quantitative interpretation, as in the example *Estupiñá se aburría algunas veces, por más que no lo declarase* (*Fortunata*: 1.257) ‘Estupiñá got bored sometimes, even if he did not say it’ (see RAE/ASALE, 2009, pp. 3620–3621).

- b. *como naturalmente fuese de flaco coraçon començo de llorar agramente e por mucho que lo consolavan los que çerca del estavan ninguna consolaçion queria oyr ni resçebir*
 'And as by his own nature he was weak of heart, he started to cry sorely, and, no matter how much those who were near him tried to console him, he did not want to hear or receive any consolation'
 (Enrique IV: 468, ca. 1481–1482)

4.2 Spanish *puesto que*: From concessive to causal?

The expected diachronic relation between causal and concessive clauses is unidirectional: from causal (simpler construction) to concessive (more complex construction), as shown in the development of *por (más) que*. Nevertheless, all the facts are not always totally consistent with this path of development. Both in Spanish and Catalan we find cases where a concessive construction later acquires a causal value and where initially ambiguous constructions end up as purely causal ones. An example of the first type is provided by Sp. *puesto que* 'since, given that', whereas Sp. *ya que*/Cat. *ja que* would be in the second group. Here we will focus on the evolution of *puesto que*, which is particularly striking.

In Spanish (and Catalan) we find several expressions consisting of a past participle and the conjunction *que*, that is, with the same structure as English *given that, provided that*: Sp. *puesto que* / Cat. *posat que* lit. 'put that', Sp. *dado que*/Cat. *donat que* 'given that', Sp. *visto que*/Cat. *vist que* 'seeing that', lit. 'seen that'. The origin of these expressions is to be found in absolute participial constructions with a subordinate clause that functions as grammatical subject of the participle. These participial expressions are generally preposed to the main clause, and their meaning is largely derived from the context. The perfective nature of the participle implies an event prior to that expressed in the main clause. From this, a causal connection can easily arise, since causes, by definition, temporally precede their effects. Depending on the context, the causal connection may be interpreted as involving an effective cause (causal), a condition (conditional), or an ineffective cause (concessive).

We illustrate these possible meanings with Old Catalan examples in (25). In (25a) *vist que* 'seeing that' is interpreted as causal, whereas in (25b) *posat que* 'provided that' has a conditional interpretation and in (25c) we have a conditional concessive interpretation of the same expression.⁵

5. For more details on Cat. *posat que*, see Rofes (2012b, section 4.12).

- (25) a. *E vist que la illa de Cerdenya era rebel·lada, lo rey En Pere hi tramès lo noble e vezcomte de Cabrera ab ·XXXX· gualeres*
 ‘And seeing that the Island of Sardinia had revolted, King Pere sent there the noble viscount of Cabrera with forty galleys’
 (Sumari: 142, 15th c.)
- b. *Bé és consonant a rahó, senyor, que ho haja oblidat, posat que ho sabés, tant gran diversitat de cogitacions ha torbat e combatut lo meu enteniment despuys que vós passàs d’esta vida*
 ‘It is well according to reason, my lord, that I should have forgotten it, provided that I knew it; such great diversity of cogitations has disturbed and battered my understanding ever since you left this life’
 (Somni: 182, 14th c.)
- c. *Mas vull-te mostrar que en les felicitats mundanals no ha bé, sinó sola ymage d’aquell e, posat que n’hi haja, no·l pot hom aconseguir*
 ‘But I want to show you that in worldly pleasures there is no good, but only an image of it, and even if there were some, one cannot obtain it’
 (Somni: 204–205, 14th c.)

In Spanish things are essentially as they are in Catalan, with the only difference that *puesto que* is first attested only in clauses with conditional concessive value ‘even if’, as in (26):

- (26) *ca puesto que la obra fuesse en sí mala, si non fuesse mal fecha nin faziéndola escogiendo que era mala, non serié del todo mala*
 ‘Because, even if the work was bad in itself, if it were not badly done and badly done on purpose, it would not be completely bad’ (Lucanor: 310, 1335)

The specific value of the construction depends, in part, on the meaning of the participle. Thus, in *vist/visto que* lit. ‘(having) seen that’, we have a causal value because what has been seen is assumed to be real. The use of the indicative mood in this construction is consistent with this value. Instead *posat/puesto que* has a conditional or conditional concessive value. This follows from the fact that *posar/poner* ‘to put’ here has the meaning of ‘to suppose, to imagine’ (Cat. *posem per cas que vinguen*/Sp. *pongamos que vienen* ‘suppose they come’), which explains the use of the subjunctive with *puesto que* at the initial stage.

The Spanish construction *puesto que* is particularly frequent among participial constructions. A well-known effect of grammaticalisation is phonological reduction. In our case we notice that in this construction *puesto* is usually unstressed (see Quilis, 1993, p. 393; RAE, 1973, p. 70), whereas other participles retain the stress in exactly the same context (we add accent marks): *visto que lo sábés* ‘seeing that you know it’, *dádo que lo sábés* ‘given the fact that you know it’, with stress on the participle, but *puesto que lo sábés* ‘since you know it’, usually without stress on the participle.

The use of *puesto que* and the more formal *puesto caso que* increases from the 14th to the 17th centuries. As is common with conditional constructions, together with its increase in use, we observe its extension to indicative contexts, as can be seen in Table 5.

Table 5. Clauses with *puesto (caso) que*

Text and year	Proposed			Postposed		
	Ind.	Subj. f	Subj. nf	Ind.	Subj. f	Subj. nf
<i>Lucanor</i> 1335			100% (1/1)			
<i>Pedro I</i> 1370	8.3% (1/12)	16.6% (2/12)	41.6% (5/12)	8.3% (1/12)		25% (3/12)
<i>Generaciones</i> 1450–1455			100% (2/2)			
<i>Celestina</i> 1499			83.3% (5/6)			16.6% (1/6)
<i>Diálogo</i> 1535				33.3% (1/3)	33.3% (1/3)	33.3% (1/3)
<i>Coloquios</i> 1547–1551			33.3% (1/3)	33.3% (1/3)	33.3% (1/3)	
<i>Bernal</i> ca. 1568	52.3% (55/105)		3.8% (4/105)	40.9% (43/105)		2.8% (3/105)
<i>Quijote I</i> 1605	31.3% (21/67)		11.9% (8/67)	53.7% (36/67)		2.9% (2/67)

The first clear examples of the locution *puesto que* are all found in subjunctive clauses. An example was given in (26) above. The subjunctive still dominates in the 15th century, but things start to change in the first part of the 16th century and in the second half of this century the indicative mood is already more frequent with this expression than the subjunctive, as illustrated in the examples in (27).

- (27) a. *en aquella sazón estaba muy adeudado y pobre, puesto que tenía buenos indios de encomienda y sacaba oro de las minas*
 ‘At that time he owed a lot of money and was very poor, even though he had good Indians of *encomienda* and extracted gold from the mines’
 (*Bernal*: 51, ca. 1568)
- b. *Así debe de ser – respondió Sancho –, puesto que yo no lo sé*
 ‘That is the way it must be – Sancho replied – even though I don’t know it’
 (*Quijote*: 187, 1605)

In the second part of the 16th century we find the first examples with causal value ('since'), as in (28a), and this causal use becomes frequent in the second half of the 17th century, as in (28b):

- (28) a. *Aquella noche dormimos con más sosiego que la pasada, puesto que teníamos mucho recaudo de corredores y espías y velas y rondas*
 'That night we slept more calmly than the night before, since we had many night guards and spies and rounds' (Bernal: 159, ca. 1568)
- b. *Y puesto que estamos solos, / señor don Álvaro, hablemos / más claramente los dos*
 'And, since we are alone, Don Alvaro, let us speak more clearly, the two of us' (Alcalde: 173, 1651)

What is unexpected here is the change in meaning from concessive ('provided that', 'even if') to causal ('since'), against the cross-linguistic tendency. We would suggest that this is not due to a process of grammaticalisation *per se*, but, rather, to a reinterpretation that is made possible by the ambiguity that is inherent to absolute constructions. In this case, the fact that *poner*, besides having the value of 'suppose' can also be interpreted as 'set down', makes a causal interpretation possible when *puesto que* is used with an indicative verb introducing a factual clause.

In fact, we find a similar reinterpretation in the case of the expression *comoquier(a) que* in the 15th century.⁶ This construction becomes specialised for the expression of pure concessives from its origins in the 13th century, even though initially it mostly appears with verbs in the subjunctive mood, as in (29a). The indicative mood becomes more common in the 14th century, as in (29b), and in the 16th century we find the first examples with a clear causal value, as in (29c)–(29d).

- (29) a. *E comoquier que ellos anden errados en la creencia, los qui la fe de Jesucristo non tienen, però muchas buenas palabras e ciertas e con razón dixieron en el fecho de la Biblia e en los otros saberes, e grandes sabios fueron e son aún oy*
 'And even though they err in their belief, those who do not have the faith of Jesus Christ, they [said] many good and true words and they spoke the truth about the Bible and in other forms of knowledge, and they were and still are very learned' (General estoria: 1.160, 1270–1280)

6. On the origin and evolution of this expression, see Vallejo (1925, p. 82), Rivarola (1976, sections 1.2.3.2 and 2.2.2.3), Bartol (1986, pp. 189–198), Montero (1992b), Cortés (1992, pp. 164–174), Herrero (2005, pp. 445–446), Ridruejo (2009) and Pérez Saldanya and Salvador (2014, Section 3.4).

- b. *como quier que era moço segund sus días, era assaz de sotil entendimiento*
 ‘Even though he was a lad by his age, he had a very subtle understanding’
 (*Lucanor*: 62, 1335)
- c. *No sé nada./Como quier que fui criada/donde siempre fui servida,/sé muy poco de colada/y menos de aquesta vida*
 ‘I don’t know anything. Since I was raised in a place where I was always served, I know very little about doing the laundry and even less about this life’
 (*Comedias*: 110, 1517)
- d. *y como quiera que sola la ratihabitación no es suficiente de causar obligación en el matrimonio sin nueva voluntad y contrato, síguese que es necesario que nuevamente se casen y consientan*
 ‘And since declaration of intent by itself is not sufficient to cause obligation in marriage without new intention and contract, it follows that it is necessary for them to marry again and consent’
 (*Lisandro*: 214–215, 1542)

Finally, a somewhat parallel development is also found in the construction *ya que* ‘since’ (< ‘as soon as’) and its Catalan counterpart *ja que*. These expressions have a more complex evolution, which we will simply outline here. This construction’s original meaning is one of immediate temporal anteriority, from which both a causal meaning (in indicative clauses) and a concessive meaning (initially in subjunctive clauses) developed. The concessive meaning lost vitality due to the ambiguity between causal and concessive interpretations in indicative clauses, and it is not normal at present.

5. Conclusions

In this paper we have considered the main properties and diachronic paths of development of concessive conjunctions in Spanish and Catalan. We have focused on two conjunctions: *aunque* and *puesto que*, and their Catalan counterparts. The conjunction *aunque* is the most basic concessive conjunction in present-day Spanish and transparently derives from *aun* ‘even’ + *que* ‘that’. The Catalan concessive conjunction *encara que* is equally transparent in its structure. These conjunctions thus illustrate the common development of concessive conjunctions from scalar adverbs, a grammaticalisation pattern attested in many languages (including English concessive expressions with *even*). We have shown, through the analysis of a historical corpus, that these expressions initially had a conditional concessive value (‘even if’), later also assumed a purely concessive value (‘although’) and, finally, also acquired a restrictive adversative value (‘but’) in certain contexts.

The evolution of *puesto que* ‘since’ is, at first glance, surprising. Nowadays it is a causal conjunction, but originally it clearly had a conditional concessive meaning ‘even if’ or ‘supposing that’. This goes against the common trend from causal to concessive meaning illustrated by other expressions such as Spanish *por más que* or Catalan *per més que*. We have argued that this change in meaning arose from the use of *puesto que* in pure concessive clauses (i. e. factual clauses with indicative), and was made possible by the polysemy of *puesto* ‘put’, finding parallels in other developments.

References

- Algeo, J. E. (1981). Constructions of the type *por (per) grande que seja* in Medieval Portuguese and Old Spanish. *Journal of Hispanic Philology*, 5, 179–184.
- Ascombre, J.-C., & Ducrot, O. (1983). *L'argumentation dans la langue*. Brussels: Pierre Margada.
- Bartol Hernández, J. A. (1986). *Oraciones consecutivas y concesivas en las Siete Partidas*. Salamanca: Acta Salmanticensia, Studia Philologica Salmanticensia.
- Brucart, J. M. (1993). Sobre la estructura de SComp en español. In A. Viana (Ed.), *Sintaxi. Teoria i perspectives* (pp. 59–102). Lleida: Publicacions de la Universitat de Lleida.
- Cortés Parazuelos, M. H. (1992). La expresión de la concesividad en español. (Unpublished doctoral dissertation). Universidad Complutense de Madrid, Madrid, Spain.
- Elvira, J. (2003). Sobre el origen de la locución concesiva *por mucho que* y similares. In *Estudios ofrecidos al profesor José Jesús de Bustos Tovar* (Vol. 1; pp. 217–231). Madrid: Editorial Complutense.
- Elvira, J. (2005). Metonimia y enriquecimiento pragmático: a propósito de *aunque*. *Dicenda*, 33, 71–84.
- Gast, V., & van der Auwera, J. (2011). Scalar additive operators in the languages of Europe. *Language*, 87, 2–54. doi:10.1353/lan.2011.0008
- Harris, M. (1988). Concessive clauses in English and Romance. In J. Haiman & S. A. Thompson (Eds.), *Clause combining in grammar and discourse* (pp. 71–99). Amsterdam: John Benjamins. doi:10.1075/tsl.18.06har
- Haspelmath, M., & König, E. (1998). Concessive conditionals in the languages of Europe. In J. van der Auwera (Ed.), *Adverbial constructions in the languages of Europe* (pp. 563–640). Berlin: Mouton de Gruyter. doi:10.1515/9783110802610.563
- Herrero Ruiz de Loizaga, F. J. (2005). *Sintaxis histórica de la oración compuesta en español*. Madrid: Gredos.
- König, E. (1985a). Where do concessives come from?: On the development of concessive connectives. In J. Fisiak (Ed.), *Historical semantics: Historical word-formation* (pp. 263–282). Berlin: Mouton de Gruyter. doi:10.1515/9783110850178.263
- König, E. (1985b). On the history of concessive connectives in English. Diachronic and synchronic evidence. *Lingua*, 66, 1–19. doi:10.1016/S0024-3841(85)90240-2
- König, E. (1988). Concessive connectives and concessive sentences: Cross-linguistic regularities and pragmatic principles. In J. A. Hawkins (Ed.), *Explaining language universals*, (pp. 145–166). London: Blackwell.
- Martín Puente, C. (2002). *Las oraciones concesivas en la prosa clásica*. Zaragoza: Universidad de Zaragoza.

- Montero Cartelle, E. (1992a). Tendencias en la expresión de la concesividad en el castellano medieval. *Verba*, 19, 107–128.
- Montero Cartelle, E. (1992b). Origen, cronología y capacidad de combinación modal de la conjunción concesiva *comoquier que*. In J. A. Bartol Hernández, J. F. García Santos & J. de Santiago Guervós (Eds.), *Estudios filológicos en homenaje a Eugenio de Bustos Tovar* (Vol. 2; pp. 657–666). Salamanca: Universidad de Salamanca.
- Montero Cartelle, E. (1992c). La trayectoria cronológica y modal de la expresión concesiva *maguer(a) que*. In M. Ariza, R. Cano, J. M. Mendoza & A. Narbona (Eds.), *Actas del II Congreso Internacional de la Lengua Española* (Vol. 1; pp. 701–710). Madrid: Pabellón de España.
- Montero Cartelle, E. (2002). La importancia del modo en la evolución de la expresión concesiva. In M. T. Echenique Elizondo & J. P. Sánchez Méndez (Eds.), *Actas del V Congreso Internacional de Historia de la Lengua Española* (Vol. 1, pp. 795–801). Madrid: Gredos.
- Pérez Saldanya, M. 1999. El modo en las subordinadas relativas y adverbiales. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 2; pp. 3253–3322). Madrid: Espasa.
- Pérez Saldanya, M., & Salvador, V. (1995). Fraseología de l'encara i processos de gramaticalització. *Caplletra*, 18, 85–108.
- Pérez Saldanya, M., & Salvador, V. (2014). Oraciones subordinadas concesivas. In C. Company Company (Ed.), *Sintaxis histórica de la lengua española. Tercera parte. Preposiciones, adverbios y conjunciones. Relaciones interoracionales* (Vol. 3; pp. 3699–3839). Ciudad de México: Fondo de Cultura Económico/Universidad Nacional Autónoma de México.
- Quer, J. (1998). *Mood at the interface*. The Hague: HAG.
- Quilis, A. (1993). *Tratado de fonología y fonética españolas*. Madrid: Gredos.
- Real Academia Española. (1973). *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- Real Academia Española/Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.
- Ridruejo, E. (2009). Sobre el proceso de gramaticalización de *comoquiera que*. In M. Veyrat & E. Serra (Eds.), *Estudios dedicados al profesor Ángel López García con ocasión de su sexagésimo aniversario* (Vol. 1; pp. 523–534). Madrid: Arco/Libros.
- Rivarola, J. L. (1976). *Las conjunciones concesivas en español medieval y clásico*. Tübingen: Max Niemeyer.
- Rivas, E. (1989). Observaciones sobre las concesivas. Su comparación con las condicionales y las adversativas. *Verba*, 16, 237–255.
- Rodríguez Molina, J., & Enrique-Arias, A. (2018). *Si* as a Q-particle in Old Spanish. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morphosyntax* (pp. 249–274). Amsterdam: John Benjamins. doi:10.1075/ihll.16.11rod
- Rodríguez Rosique, S. (2008). *Pragmática y gramática. Condicionales concesivas en español*. Frankfurt: Peter Lang.
- Rofes, X. (2007). Aportació a l'estudi de la conjunció concessiva *jatsia*. *Caplletra*, 42, 75–106.
- Rofes, X. (2010). Panorama de les construccions concessives en temps de Jaume I. In G. Colón & Ll. Gimeno (Eds.), *La llengua catalana en temps de Jaume I* (pp. 155–195). Castelló: Universitat Jaume I.
- Rofes, X. (2012a). Les construccions concessives en el *Curial e Güelfa*. In A. Ferrando (Ed.), *Estudis lingüístics i culturals sobre Curial e Güelfa* (pp. 743–794). Amsterdam: John Benjamins. doi:10.1075/ivitra.3.31rof

- Rofes, X. (2012b). Condicionals concessives i altres construccions de valor concessiu en el *Curial e Güelfa*. In A. Ferrando (Ed.), *Estudis lingüístics i culturals sobre Curial e Güelfa* (pp. 795–842). Amsterdam: John Benjamins. doi:10.1075/ivitra.3.32rof
- Salvador, V. (2010). Entre la retòrica i la gramàtica: estructures de la concessivitat en català. In K. Faluba & I. Szijj (Eds.), *Actes del Catorzè Col·loqui Internacional de Llengua i Literatura Catalanes* (Vol. 3; pp. 17–42). Barcelona: Publicacions de l'Abadia de Montserrat.
- Vallejo, J. (1922). Notas sobre la expresión concesiva. *Revista de Filología Española*, 9, 40–51.
- Vallejo, J. (1925). Sobre un aspecto estilístico en don Juan Manuel. In *Homenaje ofrecido a Menéndez Pidal: miscelánea de estudios lingüísticos, literarios e históricos* (Vol. 2; pp. 63–85). Madrid: Hernando.

Spanish corpus

- Bernal = Bernal Díaz del Castillo, *Historia verdadera de la conquista de la Nueva España*, J. A. Barbón Rodríguez (Ed.). Ciudad de México: El Colegio de México/Universidad Autónoma de México/Agencia Española de Cooperación Internacional, 2005.
- Calila* = *Calila e Dimna*, J. M. Cacho Blecua, & M. J. Lacarra (Eds). Madrid: Castalia, 1993.
- Celestina* = Fernando de Rojas, *La Celestina*, D. S. Severin (Ed.). Madrid: Cátedra, 1993.
- Cid* = *Cantar de mio Cid. Texto, gramática y vocabulario*, vol. 3: *Texto*, R. Menéndez Pidal (Ed.). Madrid: Espasa-Calpe, 1944–1945.
- Coloquios* = Pedro Mejía, *Diálogos o Coloquios*, A. Castro Díaz (Ed.). Madrid: Cátedra, 2004.
- Comedias* = Bartolomé de Torres Naharro, *Comedias*. Madrid: Castalia, 1973.
- Corbacho* = Alfonso Martínez de Toledo, *Arcipreste de Talavera o Corbacho*, J. González Muela, & M. Ciceri (Eds.). Madrid: Castalia, 1985.
- Diálogo* = Juan de Valdés, *Diálogo de la lengua*, C. Barbolani (Ed.). Madrid: Cátedra, 1990.
- Enrique IV* = *Crónica anónima de Enrique IV de Castilla*, M. P. Sánchez Parra (Ed.). Madrid: Ediciones de la Torre, 1991.
- Generaciones* = Fernán Pérez de Guzmán, *Generaciones y semblanzas*, J. A. Barrio (Ed.). Madrid: Cátedra, 1998.
- General Estoria* = Alfonso X, *General Estoria. Primera parte*, P. Sánchez Prieto-Borja (Ed.). Madrid: Biblioteca Castro.
- Hombre* = Francisco Gutiérrez de los Ríos y Córdoba, *El hombre práctico*, J. Pérez Magallón, & R. P. Sebold (Eds.). Córdoba: Publicaciones Obra Social & Caja Cultural Sur, 2000.
- Lazarillo* = *Tri-linear edition of Lazarillo de Tormes of 1554 (Burgos, Alcalá de Henares, Amberes)*, J. V. Ricapito (Ed.). Madison: The Hispanic Seminary of Medieval Studies, 1987.
- Lisandro* = Sancho de Muñón, *Tragicomedia de Lisandro y Roselia*, R. Navarro Durán (Ed.). Madrid: Cátedra, 2009.
- Lucanor* = Don Juan Manuel, *El Conde Lucanor o Libro de los enxiemplos del conde Lucanor et de Patronio*, J. M. Blecua (Ed.). Madrid: Castalia, 1984.
- Pedro I* = Pero López de Ayala, *Corónica del rey don Pedro*, C. L. Wilkins, & H. M. Wilkins (Eds.). Madison: The Hispanic Seminary of Medieval Studies, 1985.
- Proverbios* = Sem Tob de Carrión, *Proverbios Morales*, P. Díaz-Mas, & C. Mota (Eds.). Madrid: Cátedra, 1998.
- Quijote* = Miguel de Cervantes, *Don Quijote de la Mancha*, F. Rico (Ed.). Madrid: Instituto Cervantes-Crítica.

Varones = Fernando Pulgar, Hernando, *Claros varones de Castilla*, M. A. Pérez Priego (Ed.). Madrid: Cátedra, 2007.

Catalan corpus

Cocentaina = *Llibre de Cort de Justícia de Cocentaina (1269, 1275–1278, 1288–1290)*, J. Torró (Ed.). Valencia: Universitat de València, 2009.

Col·loquis = Cristòfor Despuig, *Los col·loquis de la insigne ciutat de Tortosa*, E. Duran (Ed.). Barcelona: Biblioteca Torres Amat/Curial EdicionsCatalanes, 1981.

Diàlegs = Montserrat Alegre Urgell, “*Diàlegs*” de Sant Gregori. *Retranscripció i estudi lingüístic de la versió catalana de 1340*. Ph.D dissertation, Universitat de Barcelona, 2004.

Doctrina = Ramon Llull, *Doctrina pueril*, J. Santanach i Suñol (Ed.). Palma: NEORL, 2005.

Epistolaris = *Epistolaris d’Hipòlita Roís de Liori i d’Estefania de Requesens (segle XVI)*, A. Batlle (Ed.). València: Universitat de València, 2003.

Fets = *Llibre dels fets del rei en Jaume*, J. Bruguera (Ed.). Barcelona: Barcino, 1991.

Furs = M. A. Diéguez, & J. Torruella i Casañas, *Concordança dels «Furs» de Jaume I, DBT/ Programa d’Anàlisi Textuals*. Alacant: IIFV, 2003.

Generalitat = *Dietaris de la Generalitat de Catalunya. Anys 1411 a 1539*, Ll. Cases i Loscos, J. Fernández i Trabal, & L. Pagarolas i Sabaté (Eds.). Barcelona: Generalitat de Catalunya, 1994.

Memòries = Antoni Simon i Tarrés, *Cavallers i ciutadans a la Catalunya del cinc-cents*. Barcelona: Curial, 1991.

Somni = Bernat Metge, *Lo somni*, S. Maria Cingolani (Ed.). Barcelona: Barcino, 2006.

Sumari = *Sumari d’Espanya per Berenguer de Puigpardines*, J. Iborra (Ed.). Valencia: Universitat de València, 2000.

Tirant = Joanot Martorell, *Tirant lo Blanch*, A. Hauf (Ed.). València: Tirant lo Blanch, 2005.

Vida = *Spill de la vida religiosa*, A. Bover (Ed.). Barcelona: Universitat de Barcelona, ms.

Vides = *Vides de Sants Rosselloneses*, Ch. S. Maneikis Kniazzezh, & E. J. Neugaard (Eds.). Barcelona: Fundació Salvador Vives Casajuana, 1977.

Si as a Q-particle in Old Spanish

Javier Rodríguez Molina & Andrés Enrique-Arias

Universidad de Granada / Universitat de les Illes Balears/Harvard University

This study examines the origin of Old Spanish direct interrogative *si*, a structure that has barely been described in the literature. An analysis of historical texts shows that it originates from the Classical Latin conditional conjunction *si*, which in Late Latin became a complementiser in embedded interrogatives and eventually a polar interrogative particle. This evolution is documented primarily in biblical translations. However, its continuation in Medieval Romance varieties proves that direct interrogative *si* was used in vernacular Latin. The evolution of direct interrogative *si* constitutes a case of insubordination (i.e. the conventionalised use in a main clause of a subordinate clause marker), a process that has not been considered in accounts of the historical origin of interrogative particles in grammaticalisation studies.¹

Keywords: Spanish; interrogative particle; Q-particle; *Si*; question; grammaticalisation; insubordination; Latin; conditional conjunction

1. Introduction

An aspect of Old Spanish syntax that has passed virtually unnoticed in the literature is the presence of interrogative structures introduced by the particle *si*. As illustrated in the examples in (1), in Old Spanish *si* could be used optionally

1. Javier Rodríguez Molina and Andrés Enrique-Arias have benefited from two separate grants from the Spanish Ministry of Science and Innovation (AEI/FEDER, UE), references FFI2015-64722-P and FFI2014-59135-R, respectively, and a grant from the BBVA Foundation Program for Scientific Research Teams/Digital Humanities 2016. Andrés Enrique-Arias has also received funding from the Ministry of Education, Culture and Sports of Spain for a research stay at Harvard University (reference PRX17/00505, 2017–18). Likewise the authors are grateful to Miriam Bouzouita, Ioanna Sitaridou, Luis Girón Negrón and audiences in Madrid, Heidelberg, Tübingen, Salamanca and Cambridge (Massachusetts) for helpful comments on previous versions of this paper.

to introduce a polar interrogative (1a), or an alternative interrogative (1b). These lesser known structures occur next to the widely attested and studied *si* that introduces indirect interrogatives (1c).

- (1) a. *Dixo el nuestro Señor: ¿Si me celaré de Abraam de lo que quiero fer?*
 ‘Then our Lord said: “Shall I hide from Abraham what I am about to do?”’
 (BM, Fazienda, Gen 18:17, ca. 1200)
- b. *Díxole: ¿Si eres el mi Jasón o si eres alguna vesión que en sueños me has aparecido [...]?*
 ‘He said to him: “Are you my Jason or are you some vision that has appeared in my dreams?”’
 (Bienandanzas, 41r, 1471–1476)
- c. *Enbióle dezir si quería casar con una su parienta*
 ‘He sent someone to ask if he wanted to marry a relative of him’
 (Atalaya, 22r, 1443–1454)

Interrogative *si* in direct questions is already attested in some of the earliest written manifestations of Spanish – as far back as the early 13th century – and has a significant presence across medieval texts. This structure is still at work during the 16th and 17th centuries. However, from the 18th century on, direct interrogative *si* appears rather infrequently and only in rhetorical questions. In modern Spanish, direct questions introduced by *si* are very infrequent: they sound old-fashioned and are restricted to written texts (Escandell, 1999, p. 3968).

As already mentioned, direct interrogative *si* has barely been noticed by linguists and no detailed historical analysis of its meaning and functions has been formulated. We aim to fill this gap by tackling a number of questions relating to the origin and evolution of this particle: what is the historical origin of the direct interrogative *si*? What nuances of meaning did it develop? How frequently and in what kind of texts did it appear? Are there similar structures in other Romance languages? Our aim is to establish the chronology for this interrogative particle (Q-particle) and to determine its grammatical properties in relation to its Latin ancestors and its Romance cognates (cf. Wright, on the importance of Early Romance data).

In order to track direct interrogative *si* structures through the history of Spanish, we have carried out searches in literary works from different historical periods. As direct interrogative *si* is particularly frequent in medieval biblical translations, we have made extensive use of the *Biblia Medieval* corpus, an online database that enables users to view in parallel columns the extant Medieval Spanish versions of the Bible and compare them side-by-side with their Latin and Hebrew sources.

This paper is organised as follows: first, we outline some notions that will be relevant in our discussion of the origin and development of interrogative *si* in the

history of Spanish, namely the classification of the different types of interrogatives and a summary of how they are expressed in the major Romance languages. Then, we trace the changes undergone in the evolution of interrogative structures from Latin to Old Spanish and the other major Romance languages with special attention to the presence of cognates of interrogative *si*. Finally, we provide an overview of the grammaticalisation patterns of Q-particles from a typological perspective, in order to shed light on the origin and evolution of the Spanish Q-particle *si*.

2. Interrogative constructions in Romance

Interrogative constructions are classified according to the information they request as Polar Interrogatives (or yes/no-Qs), Content Interrogatives (or *wh*-Qs) and Alternative Interrogatives (or alt-Qs) (Köning & Siemund, 2007; Siemund, 2001). We will leave aside content interrogatives, as they are not relevant to our discussion of interrogative *si*.² Yes/no-Qs can be defined as interrogative structures whose expected answer is the equivalent of “yes” or “no”, because they inquire about the truth or falsity of the proposition they express, as in (2a). Yes/no-Qs can be stated in a positive or negative way, as in (2a)–(2b): while the former is neutral and does not bias the answer towards any of the two alternatives, the latter usually implies that the speaker expects that the answer will be a negative one.³ Alt-Qs inquire which element of a set of alternatives makes an open sentence true (2c).

- (2) a. *Do you like beer?*
 b. *You don't like beer?*
 c. *Would you like tea or coffee?*

Interrogative constructions may be found both in main clauses and embedded structures. In most cases languages do not employ the same coding strategies for both domains: for instance, English uses an interrogative auxiliary in main clauses, as in (3a), while embedded interrogative structures are marked by the conditional conjunction *if* (cf. (3b)). In contrast, in some languages, such as Classical Latin,

2. Content Interrogatives request that an answer specify the value of the variable bound by the *wh*-word (*what, where, when, who*, etc.).

3. Negative yes/no-Qs allow for certain ambiguity, since a positive answer to an utterance like (2b) could mean either “yes, I don't like beer” or “yes, I do like beer”, depending on whether the speaker has in mind the truth-value of the situation or the polarity used in the question (RAE/ASALE, 2009, § 42.10).

the same Q-particle may be used to code matrix and embedded interrogatives (cf. (3c)–(3d)).

- (3) a. *Do you like coffee?*
 b. *He is asking you if you like coffee*
 c. *Eloquar an sileam?*
 ‘Shall I speak up or shall I be quiet?’ (Verg. Aen., 3.39)
 d. *Haud scio an congrediar*
 ‘I don’t know if I will come’ (Plaut., Epid., 543)

Though interrogatives are conventionally associated with the speech act of requesting information, this is not always the case. Interrogative constructions that do not code an interrogative speech act, and in which, therefore, there is no expectation of a reply, are called non-canonical interrogatives (Siemund, 2001, pp. 1026–1027). Rhetorical questions (RQs), which can be found in practically every language, constitute a well-known type of non-canonical interrogative structure. In RQs the speaker is not requesting information but rather intends to convey a reinforced statement that entails a reversal of polarity ($\text{?}p? = \neg p$; $\text{?}\neg p? = p$). For instance, when asking a question like *who wants to live forever?* the speaker is making a statement (i.e. ‘nobody wants to live forever’) rather than a genuine request for information.⁴

While natural languages exhibit a wide range of variation in the strategies employed for coding interrogative constructions (Dryer, 2011; Harris & Campbell, 1995, pp. 294–298; Köning & Siemund, 2007; Siemund, 2001), standard Romance languages use a rather narrow set of options, which may be summarised as follows (Arnaiz, 1998, pp. 56–60):⁵

4. There are at least four types of non-canonical interrogative sentences besides RQs: surprise interrogatives, “can’t find the value”-interrogatives, exclamative interrogatives and imperative interrogatives (Obenauer, 2004). In this study we will focus on RQs because, of all the non-canonical types, they are the ones most commonly represented in historical texts.

5. For a detailed account of interrogative structures in Romance, see the following: Escandell (1999, 2012) and RAE/ASALE (2009, §§ 42.6, 43.7, 43.8) for Spanish; Elsig (2009) for French; Wheeler, Yates & Dols (1999, § 27.1) and Payrató (2002, pp. 1201–1212) for Catalan; Fava (1995, pp. 70–126), Maiden and Robustelli (2007, pp. 143–148) for Italian; Mira Mateus et al. (2006, pp. 460–479) for Portuguese. Old Spanish interrogative structures have never been described in detail, as opposed to Old French, for which we do have some detailed descriptions (see Nyrop, 1930, §§ 382–399; Posner, 1995; Buridant, 2000, pp. 682–694 among others).

1. Intonation is by far the most widespread coding strategy for deriving interrogatives.
2. Interrogative structures may entail word order variation, usually displaying a verb-subject inversion pattern, derived by raising the tensed verb to the CP layer.
3. *Wh*-words in content interrogatives surface at the sentence beginning position (however, Brazilian Portuguese, colloquial French and some Italian dialects allow *wh-in situ* in main clauses).
4. Q-particles are scarcely used, leaving aside French (postverbal attachment of Q-particle *-t*, sentence-initial Q-particle *est-ce que*) and Portuguese (*é que*).
5. Polar embedded interrogatives are introduced by *si/se* and require neither a particular word order nor a different intonation pattern.

These general tendencies are summarised in Table 1.

Table 1. Romance languages interrogative strategies

Language	Polar interrogatives			Polar embedded
	Intonation	Inversion	Q-particle	Complementiser
Portuguese	yes	yes	<i>é que</i>	<i>se</i>
Spanish	yes	yes	∅	<i>si</i>
French	yes	yes	<i>est-ce que, -t</i>	<i>si</i>
Italian	yes	yes	∅	<i>se</i>

In Spanish, yes/no-Qs and alt-Qs are formed by just two strategies: (a) a rising intonation pattern with the basic declarative word order (4a); or (b) optional inversion of subject and verb (4b). As for embedded interrogatives, all major Romance languages employ a reflex of Latin conditional marker *SI* (4c and 1c) (cf. Pérez Saldanya & Hualde, in this volume, for the semantic development of Spanish and Catalan conditional concessive subordinators *aunque/encara que* ‘even if’).

- (4) a. *¡Juan viene al cine?*
‘Is John coming to the theater?’
- b. *¿Viene Juan al cine?*
‘Is John coming to the theater?’
- c. *Pregunto si Juan viene al cine*
‘I wonder if John is coming to the theater’

In sum, Romance languages do not make extensive use of interrogative particles, as intonation and verb inversion are the most common strategies for forming polar interrogatives.

3. The evolution of interrogative structures: From Latin to Romance

Unlike the Romance languages, Classical Latin had a rich system of interrogative particles. Yes/no-Qs were typically marked with the enclitic particle -NE or with particles NE, NONNE, NUM or AN. Alt-Qs were signalled with AN, -NE... AN, UTRUM...AN, UTRUM ... -NE, ... AN, AN...NECNE. Indirect questions, which required the verb in subjunctive mood, were introduced by NUM, -NE, NONNE or AN (Ernout & Thomas, 1953, §§ 316–320; Bodelot, 1987, 1990). In contrast, *wh*-Qs were expressed with interrogative pronouns or adverbs, as in the Romance languages.

In its transition to Late Latin the interrogative particle system in Classical Latin underwent a series of substantial changes that can be summarised in five processes (Herman, 1996; Brown et al., 2009, pp. 513–514, 520):

1. Gradual disappearance of -NE and all other interrogative particles (AN, UTRUM, NONNE, etc.) between the 1st and 4th centuries.
2. Simultaneously, replacement of NUM by NUMQUID and NONNE by NUMQUID NON (NE) in polar interrogatives and in indirect questions.
3. Use of the conjunction AUT ‘or’ to mark alternative interrogatives (the use of AUT and ET to join two interrogative clauses was already possible in Classical Latin) (Ernout & Thomas, 1953, § 429).
4. Indicative becomes the usual mood in indirect questions, replacing the subjunctive.
5. NUM and NUMQUID are replaced by SI in indirect questions, a structure that would eventually prevail in all Medieval Romance languages (Spanish and French *si*, Portuguese and Italian *se*).⁶

This last change is particularly important to understand the historical origin of Spanish Q-particle *si*. In Classical Latin SI was a conditional conjunction (5a), which expanded its functions to become a particle to mark indirect questions (5b)–(5c). According to Ernout and Thomas (1953, p. 320), this last function originated in Archaic Latin, in sentences formed with speech and cognition verbs such as UIDEO ‘see’ and continued in Classical Latin but was still restricted to very few verbs (Bodelot, 1987, 1990).⁷ The increasing use of this structure at the expense of

6. Portuguese and Italian *se* must have derived from a form *SĪ besides SĪ.

7. Conditional SĪ derives from the Indo European particle **sei*, from which the adverb SĪC originates as well (Ernout & Meillet, 1959, p. 622). In origin it did not necessarily have a conditional meaning, nor was it a subordinating conjunction: correlations such as *Meam rem non cures, si recte facias* ‘You wouldn’t meddle with my business, if you behaved decently’ [Pl.

NUM and NUMQUID will not become noticeable, however, until the 4th century C. E. (Herman, 1996, pp. 297–298).

- (5) a. *Si bellum omittimus, pace numquam fruemur*
 ‘If we avoid war, we will never enjoy peace’ (Cic. Ph. 7, 19)
- b. *primus ab iis quaesivit si aquam hominibus iumentisque in totidem dies quot frumentum imposuissent*
 ‘He first asked them if they had put on board water for men and beasts for just as many days as they had grain’ (Liv., 29, 25, 8)
- c. *vide si hoc utibile magis atque in rem deputas*
 ‘See if you consider this more practical, more advisable’ (Pl, Tri. 748)

The progressive encroachment of *SI* in embedded interrogatives followed two paths: on the one hand, spoken Latin favoured this form over NUM and NUMQUID (6a)–(6b) (in fact these forms have not left any Romance descendants). At the same time *SI* was adopted in the Latin used by Bible translators and Christian writers: in the Gospels we find numerous examples of *si* used to render *ei* in the Greek original (6c); we also find *si* in the Vulgate to render Hebrew Q-particles *ha* (6d) and *im* (6e) in indirect questions after verbs conveying notions such as ‘inquire’, ‘doubt’, ‘examine’ and the like. It should be noted that Greek embedded interrogatives required the indicative mood, a factor that may have favoured the tendency towards abandoning the subjunctive in these structures in Latin (Ernout & Thomas, 1953, § 321; Blaise, 1955, pp. 153–154; Herman, 1996, p. 297).

- (6) a. *Expecto si quid dicas*
 ‘I’m awaiting what you have to say’ (Pl. Trin. 98)
- b. *Quaerite igitur si uera est ista divinitas Christi*
 ‘Ask, then, whether that deity of Christ be true’ (Tert. Apol. 21:30)
- c. *dicas mihi si tu es Christus*
 ‘Tell me if you are the Messiah’ (Latin Vulgate, Mat 26:63)
- d. *uidete si est dolor sicut dolor meus*
 ‘See if there is any sorrow like my sorrow’ (Latin Vulgate, Lam. 1:12)
- e. *videamus si floruit vinea, si flores fructus parturiunt, si floruerunt mala punica*
 ‘Let us see if the vines have budded, if their blossoms have opened, if the pomegranates are in bloom’ (Latin Vulgate, Song of Salomon 7:13)

Cap. 632]) are not interpretable as a conditional structure (Ernout & Thomas, 1953, § 369; Ernout & Meillet, 1959, p. 622).

Finally, in Late Latin we find *si* used as a Q-particle to introduce direct yes/no-Qs.⁸ This new use is particularly frequent in the Vulgate and in the writings of the Fathers of the Church as illustrated in the following examples:⁹

- (7) a. *dicentes: si licet sabbatis curare?*
 ‘Saying: is it lawful to heal on the Sabbath?’ (Mat. 12:10, Blaise, 1954, p. 757)
- b. *si trades eos in manu Israhel?*
 ‘Will you deliver them into the hand of Israel?’ (*Latin Vulgate*, Sam1 14:37)
- c. *et si ad te pertinent?*
 ‘Who are these with you?’ (*Latin Vulgate*, Gen. 33:5)
- d. *interrogabant eum dicentes: Domine, si in tempore hoc restitues regnum Israhel?*
 ‘They would ask him: Lord, are you at this time going to restore the kingdom to Israel?’ (*Latin Vulgate*, Act. 1:6)

The extension of conditional SI to introduce embedded interrogatives first and eventually become a Q-particle in direct questions was favoured by a number of functional motivations (Herman, 1996, pp. 303–306). The grammaticalisation of *si* as a conjunction that introduces embedded interrogatives helped compensate for the loss of the Q-particles of Classical Latin, which were phonologically weak (i.e. enclitic -NE), or were liable to become homophonous with other words (NUM/NON). Furthermore, there was a structural parallel in the combination of QUOD, QUIA and SI with certain verbs (*scire volo si pater advenit* ‘I want to know if father will come’/*dic si pater advenit* ‘tell me if father will come’/*scio quod pater advenit* ‘I know that father will come’/*dic quod pater advenit* ‘tell me that father will come’); this structural parallel quite likely favoured the reanalysis of *si* as a subordinating conjunction and facilitated its extension to use in indirect interrogatives with all kinds of verbs. Finally, another factor to consider is the conceptual closeness between certain conditional and interrogative contexts, as they are both in the realm of the *irrealis* modality.

8. The use of SI as an interrogative particle was first described in the 19th century and is mentioned in well-known works such as Löfsted (1911, p. 327), Ernout and Thomas (1953, § 321), Hofmann-Szantyr (1965, p. 464) and Väänänen (1981, § 348); likewise there is a detailed study by Herman (1996), as well as abundant documentation in Blaise (1954, p. 757). In contrast, no Romanist has studied the continuation of this structure in the Romance languages.

9. The Vulgate uses *si* in direct interrogatives in a number of cases; in addition to the examples given in (7) cf. Sam1 23:12, Sam 2 5:19, Matthew 19:3, Luke 13:23, Joel 1:2, 2 Maccabees 7:7 among others).

Alongside the functional factors just mentioned, there are reasons to think that the extension of conditional *si* to interrogative structures was further encouraged by the presence of the conditional conjunction to introduce both indirect and, albeit with restrictions, direct interrogatives in biblical Hebrew, as well as in the Greek used in the New Testament and Deuterocanonical books of the Old Testament. In biblical Hebrew yes/no-Qs do not need to be introduced by an overt marker. As a rule, however, yes/no-Qs are coded with the Q-particle *ha*, which stands before the first word in the clause¹⁰ (8a). Disjunctive questions are usually introduced by *ha* in the first clause and *im* in the second (8b), with the latter particle deriving from the conditional conjunction *im* ‘if’. In some rare cases, a direct question is introduced by *im* (8c), but this is in principle due to the suppression of the first member of a double question.¹¹

- (8) a. *ha-šomer aḥi anokhi*
 ‘Am I my brother’s keeper?’ (Hebrew Bible, Gen. 4:9)
- b. *ha-malokh timlokh alenu im-mašol timšol*
 ‘Are you to reign over us? Are you to have dominion over us?’
 (Hebrew Bible, Gen. 37:8)
- c. *im-koah avanim koḥi im-besari nahuš*
 ‘Is my strength the strength of stones, or is my flesh bronze?’
 (Hebrew Bible, Job, 6:12)

As for Greek, while the use of *ei* in indirect interrogatives is attested early on (cf. (9a)), direct interrogative *ei* is a later development (9c), which Blass and Debrunner (1982, pp. 530–531) attribute to a Hebrew calque in the translation of the Septuagint (9b):

- (9) a. *safa d’ouk oid ei theos estin*
 ‘Yet I know not surely if he be not a god’ (Homer, Iliad 5.183)
- b. *im meēt adoni ha-melekh niyah ha-davar ha-zeh*
- c. *ei dia tou curiou mou tou basileos jeyonen to rhe:ma touto?*
 ‘Is this thing done by my lord, the King [...]?’ (1Kings 1: 27)

In turn, it has been suggested that Latin interrogative *si* in direct questions is a calque of Greek *ei* (Brown et al., 2009, p. 520), and the same could be said of

10. There are 746 cases of interrogative *ha* in the Hebrew Bible according to Clines (1993–2012, s.v. *ha*).

11. For instance, in 1 Kings 1:27, Isaiah 29:16, Job 6:12, Job 39:13.

the use of *si* in the Latin Vulgate to render Q-particles *ha* and *im* in the Hebrew original.¹²

- (10) a. *ha' umnam 'elem šedec teda'erún mešarim*
 b. *ei ale:tho:s ara dikaisune:n laleite eutheia?*
 c. *si vere utique iustitiam loquimini recta?*
 'Do you indeed decree what is right?' (Hebrew Bible, Psalms 58:2)

It is worth pointing out that the extension of the *si* particle from embedded clauses to direct interrogatives takes place in the same texts (biblical translation, religious texts) that favoured the use of *si* in embedded interrogatives. In such texts the underlying Greek and Hebrew models provide a template that favours the use of *si* in both direct and indirect interrogatives: as we have just seen, Greek particle *ei* and Hebrew particles *ha* and *im* introduce both kinds of interrogatives. In turn, both *ei* and *im* are the conjunctions that introduce conditional sentences. This means that Latin *si* is the closest equivalent to these particles and, thus, an obvious candidate to assume their functions in biblical translation and related religious texts.

While biblical language may have played a role in encouraging the adoption of *si* in yes/no-Qs we should also take into consideration that there are cases of interrogative *si* in passages where the Greek original has no instances of *ei* (cf. 11a). Moreover, other Latin texts that are not translations from Greek or Hebrew, such as the *Peregrinatio*, exhibit uses of *si* in direct questions (11b).

- (11) a. *Et respondens Iesus dixit ad legisperitos et pharisaeos*
dicens: Si licet sabbato curare? kai apokritheis ho iesous eipen [...]
lego:n eksestin to: sabbato: therapeusai ei: ou? (Vulg. Luc. 14, 3, cited by
 Herman, 1996, p. 299)
- b. *et sic singulariter interrogat episcopus uicinos eius, qui intrauit dicens: si*
bonae uitae est hic, si parentibus deferet, si ebriacus non est aut uanus?
 'As they come in one by one the bishop asks their neighbours questions
 about them: Is this person leading a good life? Does he respect his
 parents? Is he a drunkard or a boaster?'
 (Itin. Eger. 45, 3, Bertocchi & Maraldi, 2011, p. 110)

Herman (1996, p. 302) proposes that the use of *si* to introduce direct questions is a case of aborted grammaticalisation which does not survive beyond late Latin ("il n'y aura dans ces langues [romanes] aucune trace de *si* interrogatif direct"), an assertion which in light of the data presented in this paper cannot, however, be

12. Greek and Latin contacts and their mutual influence, particularly in the frame of the spread of Christianity and the translation of the Gospels from Greek into Latin, has been a prominent subject in the literature.

sustained. The existence of instances of direct interrogative *si* not linked to biblical translation coupled with the survival of the structure in Medieval Spanish as well as other Romance languages indicates that it must have had some use in colloquial speech.¹³

Table 2 below summarises the main changes in the evolution of Latin interrogative strategies.¹⁴

Table 2. Evolution of Latin interrogative strategies

LATIN	NOT EMBEDDED				EMBEDDED		IF- CLAUSE
	YES / NO-Q			ALT-Q	PARTICLE	MOOD	
	TRUTH VALUE OF THE PROPOSITION						
	HIGH	←—————→		LOW			
	ORIENTATION						
	(+)	NEUTRAL	(-)				
CL	<i>num</i>	<i>-ne</i>	<i>nonne</i>	<i>an</i>	<i>num, an, nonne</i>	SUBJ	<i>si</i>
VL	<i>numquid</i>	<i>numquid</i>	<i>numquid non</i>	<i>numquid</i>	<i>numquid</i>	SUBJ	<i>si</i>
LL	∅ (<i>si</i>)			<i>(aut)... aut</i> <i>(si)... (si)</i> <i>aut</i>	<i>si</i>	IND	<i>si</i>

4. Interrogative *si* in the history of Spanish

As already explained, Modern Spanish yes/no-Qs and alt-Qs are formed by just two strategies: rising intonation pattern with the basic declarative word order, and optional inversion of subject and verb. Historical linguists have tended to assume that Old Spanish exhibited the same devices for forming interrogative structures,

13. However, direct interrogative *si* does not appear in texts with abundant dialogue, such as the Acts of the Martyrs or the transcription of the Donatist Processes. Likewise, it is not documented in non-biblical Greek, and we find no trace of it in Latin documentation of the Merovingian era (Herman, 1996, pp. 301–302).

14. CL stands for Classical Latin, VL for Vulgar (i.e. colloquial) Latin and LL for Late Latin.

and thus the existence of Q-particle *si* has been largely ignored.¹⁵ As a result, there are no historical studies of this structure beyond very succinct observations generally restricted to its use in biblical texts.¹⁶

Although not very frequent, interrogative *si* in direct questions has a continuous presence in the history of Spanish. The structure is already attested in the *Cantar de Mio Cid*, one of the earliest literary works in Spanish (12a), and appears in a few other texts in the 13th–15th centuries. The following examples illustrate the use of *si* in both canonical (12a), (12c) and rhetorical questions (12b), (12d)–(12e), the former being typically inserted in dialogues and the latter in argumentative passages.

- (12) a. *¿Commo son las saludes de Alfons mio señor?./¿si es pagado o reçibió el don?*
‘What greetings are there from Alfonso, my lord?/is he pleased or did he accept the gift?’ (Cid, v. 1922 ca. 1200)
- b. *¿Si fue nunca qui son tan dulce oyesse en logar del mundo?*
‘Has anybody ever heard such a beautiful sound?’ (GEI, ca. 1270)
- c. *Dixieron ellos: ¿si tú viste de tiempo acá un ome estraño que avía nombre Plácidas con su mugier e con sus fijos?*
‘They said: “Have you seen lately a strange man named Placidus with his wife and children?”’ (Plácidas, 14th c.)
- d. *¿por qué non pedides la cosa certera? Ella diz: ¡Maguera! ¿Si me será dada?*
‘Why don’t you ask for a certain thing? She says: “What! Will it be given to me?”’ (Buen amor, SG1034, 14th c.)
- e. *¡Yuy, y qué milagro atán grande! ¿Si vimos nunca tal?*
‘Alas, what a great miracle! Did we ever see such a thing?’ (Corbacho, 1469)

15. Besides the use of *si*, there are other aspects of Old Spanish interrogatives that have not been studied yet, such as the productivity of subject-verb inversion and the use of other interrogative particles such as *o* ‘or’ and *e* ‘and’.

16. To our knowledge, the first mention of Old Spanish polar interrogative *si* appears in Sachs (1948–1949, p. 221), who simply observes its existence in biblical translations. A few other studies centered on the language of Old Spanish Bibles contain brief notes on this structure: Sephiha (1989), Sanchis Calvo (1991: 544), Enrique-Arias (2006, pp. 253–254, 2008, pp. 114–115, pp. 117–118), Bouzouita (2008, pp. 225–227, 2011), Almeida (2013, pp. 19–20), while García de Diego (1951, p. 401) points out its existence in 16th century literary prose. Enrique-Arias and Burguera’s (2010) study is the only in-depth analysis of the structure but is confined to the analysis of the expression of RQs in biblical translation.

- f. *Pues las cortes de los reyes, príncipes y grandes señores, ¿si hay en ellas algunos de estos deseos malditos? Dudar en ello sería pecado*
 ‘So, in the courts of kings, princes and great lords, are there any of these wicked desires? It would be sinful to doubt it’ (Corbacho, 1469)

Direct interrogative *si* continues to appear in literary works in the 16th and 17th centuries; it is featured in dialogues in narrative texts and also in plays by authors that aim to emulate real oral interactions, which indicates that the structure must have had some sort of existence in oral registers.

- (13) a. *El perro ladra, ¿si viene este diablo de vieja?*
 ‘The dog is barking, is this devilled old lady coming?’
 (Celestina, 7.213, 1501)
- b. *¿Si han muerto ya a nuestro amo?*
 ‘Have they killed our master already?’ (Celestina, 12.268, 1501)
- c. *O señora y amor mío Melibea, ¿qué piensas agora? ¿Si duermes o estás despierta? ¿Si piensas en mí o en otro? ¿Si estás levantada o acostada?*
 ‘O my lady and my beloved Melibea, what are you thinking now? Are you sleeping or are you awake? Are you thinking about me or somebody else? Are you up or sleeping?’ (Celestina, 13.280, 1501)
- d. *–Por ventura –dixo ella–, ¿si será el hijo del rey Perión de Gaula? –No sé – dixo el Rey*
 ‘By chance – she said – could he be the son of king Perion of Gaulle? –I don’t know – the king said’ (Amadís, 1.390, 1508)
- e. *¿Si es amasado de manos limpias?*
 ‘Has it been kneaded with clean hands?’ (Lazarillo, 48, 1554)
- f. *¿Si se combaten aquellos?*
 ‘Are they fighting each other?’ (Cervantes, Celos, 1, ca. 1585)
- g. *¿Si será posible que he ya hallado lugar?*
 ‘How is it possible that I have already found a place?’
 (Quijote, I, 28, 1605)
- h. *–¿Si es ella? –Tenlo por cierto*
 ‘–Is it her? –You can bet on it’ (Peribáñez, 82, 1614)

After the 1700s the direct interrogative *si* becomes very rare and is used only in non-canonical uses, such as exclamative, dubitative and rhetorical interrogatives. In the few examples that we have located the main verb in the clause is either in the future or future perfect, which in Spanish conveys a sense of doubt or supposition equivalent to ‘I wonder whether [...]’. In contrast, the direct interrogative *si* in Old Spanish could combine with verbs in any tense (cf. present in (13a), present perfect in (13b) or even past tense in (12c), (12e)). These non-canonical

structures are the only use of direct interrogative *si* that has prevailed to this day (RAE/ASALE, 2009).

- (14) a. *¿Si querrá decir que Judas [...] hizo colación con doce mil dracmas de plata?*
 ‘Does he mean that Judah made a feast with twelve thousand silver drachmas?’ (Isla, 795, 1758)
- b. *¿Si habré yo visto visiones?*
 ‘Did I see a vision?’ (Regenta 1.363, 1884)
- c. – *¡Orgullosa! ¿si creerá que no tenemos ojos [...]?*
 ‘Arrogant! Does she think we can’t see?’ (Regenta 2.80, 1885)
- d. *¿Si habré conocido en mi infancia a alguien que tenga criados sin saberlo?*
 ‘Did I meet as a child someone that had servants without knowing it?’ (Zalacaín, 208, 1908)

While direct interrogative *si* is fairly uncommon in literary texts, Old Spanish Bible translations employ it rather frequently to render the Hebrew Q-particles *ha* and *im* (or their Late Latin counterparts NUM, NUMQUID, and more rarely SI).¹⁷ *La Fazienda de Ultramar*, possibly the oldest prose work in Spanish (produced ca. 1200), contains at least ten examples of direct interrogative *si* in biblical passages translated from Hebrew (cf. Sephiha, 1989, p. 323), as illustrated in (15a). There are two more 13th century biblical translations, both from Latin. The oldest one, known as the Pre-Alphonsine Bible (ca. 1250), tends not to use any explicit markers in interrogatives. However, it does feature a few examples of the structure in question (cf. (15b)). The *General estoria* (ca. 1270–80), a historiographic work that contains numerous biblical passages translated from Latin; also features direct interrogative *si* (15c), although the most common use is *si non*.¹⁸

17. There are over a dozen manuscripts containing biblical translations into Old Spanish. In addition, there are loose books and small fragments as well as biblical passages of varying length inserted in historical works. For the 13th century texts, there are three substantial translations (one from Hebrew and two from Latin) while for texts from the 15th century, there are six different translations, all of them from Hebrew with the exception of a small number of books translated from Latin (cf. Pueyo Mena & Enrique-Arias, 2013). For the most complete information on these manuscripts and an exhaustive bibliography, see the *Biblia Medieval* project website.

18. This rather idiosyncratic use may be related to the common presence of negative words to code biased interrogatives in which a positive answer is expected (Latin *nonne*, French *n'est ce pas?*, Spanish *¿No es verdad...?* (Gordis, 1933). The same is true of Old Spanish: *¿Non sabiés cómo avía yo de entrar a esta tierra e destruirla [...]?* [GE4, ca. 1280] ‘didn’t you know how I was to enter this land and destroy it?’

- (15) a. *¿Si fuesse trobado tal omne com est?*
‘Could we find anyone else like this?’
(*BM-Fazienda*, Gen 41:38 ca. 1200)
- b. *e preguntavan: ¿Si conviene sanar en los sábados?*
‘And they questioned him, asking: “Is it lawful to heal on the Sabbath?”’
(*BM-E6*, Matt. 12:10, ca. 1250)
- c. *¿Si Efraím mio fijo sea de onrrar, si sea moço criado a vicio?*
‘Is Ephraim my son of honour, is he a child raised for sin?’ (*BM-GE*, Jer 31:20 ca. 1280)

Use of direct interrogative *si* intensifies in the translations produced in the 15th century, most of them translated from Hebrew, to the point of becoming in some translations almost an automatic calque (16a). It is used in translations from Latin as well (16b), albeit only sporadically.

- (16) a. *E díxoles: -¿Si conocedes a Labán fijo de Nahor? E dixeron: -Conocemos*
‘And he said to them: “-Do you know Laban the son of Nahor?” And they said: “-We know him”’
(*BM-E3*, Gen 29: 5, ca. 1420)
- b. *preguntóle diziente ¿Si conviene en los sábados curar?*
‘And they questioned him, asking: “Is it lawful to heal on the Sabbath?”’
(*BM-Santillana*, Matt. 12:10, ca. 1450)

The medieval translations of the Bible into Old Spanish provide hundreds of examples of direct interrogative *si* next to their Latin or Hebrew sources and side-by-side with the Old Spanish equivalents in the other versions. Thanks to the parallel nature of the *Biblia Medieval* corpus it is very easy to appreciate the range of variation in the forms used to code RQs, as illustrated in (17).

- (17) (Job 8:3)
- a. (*BM-E8*) *¿E Dios, tienes que engaña el juizio?* (ca. 1250)
- b. (*BM-GE*) *¿Si non derriba Dios el tu juizio?* (ca. 1280)
- c. (*BM-Ayala*) *¿Si él tuerce juizio?* (ca. 1407)
- d. (*BM-E3*) *¿Si Dios atuerce el juizio?* (ca. 1420)
- e. (*BM-Santillana*) *¿Si Dios atuerce tu juizio?* (ca. 1420)
- f. (*BM-E5*) *¿Quiçá Dios estuerce el derecho?* (ca. 1420)
- g. (*BM-Arragel*) *Nunca el Señor Dios atorció juicio* (1431)
‘Does God pervert justice?’

The comparison of parallel versions shows how translators use *si* next to other competing strategies such as verbs of thought (E6 *¿tienes que...?* ‘do you believe that...?’) or epistemic operators of doubt (E5 *quiçá* ‘perhaps’). Also, in some cases, given that RQs are actually conveying a statement, translators do not use an interrogative structure but a statement instead with change of

polarity (cf. Arragel translation in (17g) with negative polarity: ‘the Lord never perverted justice’).

Enrique-Arias and Burguera (2010) selected 52 examples of RQs in the Hebrew Bible (41 coded with *ha* and 11 with *im*) along with their Latin equivalents (chiefly NUM and NUMQUID) and analysed how they were rendered in Old Spanish. The study shows that, overall, direct interrogative *si* was the most frequent mechanism to convey the Q-particles in the Hebrew or Latin original (51.1% or 160/313). Furthermore, after considering only overt markers and discarding the 40 cases in which translators did not use an explicit marker to formalise the interrogative, or the 63 instances in which they did not translate using a question but a statement, then *si* is used over 76.2% (160/210), an overwhelming majority, followed at great distance by all other options: epistemic markers of doubt (*por ventura* ‘by chance’, *quicá* ‘perhaps’ 8.6%), exclamatives (*¿cómo* [...]! ‘[how [...]!’ 5.7%), connective words (*pues* ‘so’, *pero* ‘but’ 3.8%), and a few others.

The analysis of Old Spanish renderings of the Bible also suggests that medieval translators made a connection between polar interrogative *si* and the conjunction *si* that introduces embedded questions. In some cases, we find that the translator has added a matrix verb that was not present in the original in order to turn the direct question into a subordinate clause:

- (18) (Job 7:12)
- a. *ha-yam 'aní 'im tanín ki tasim 'alay mišmar*
‘Am I the sea or the dragon, that you set a guard over me?’
(Hebrew Bible)
 - b. [*Querría yo saber*] *si so yo mar o drago que me cercas con cárcel*
‘[I would like to know whether] I am the sea or the dragon, that you set a guard over me’
(BM-Arragel,)
- (19) (Num 22:38)
- a. *ha-yakhol 'ukhal dabber me'uma*
‘Do I have power to just say anything?’
(Hebrew Bible)
 - b. [*Dubdo*] *si podrá hablar en cosa alguna*
‘[I doubt whether] I have the power to say anything’
(BM-Arragel)

It is also worth pointing out that at least in some cases direct interrogative *si* is used in passages (20b)–(20c) in which the Hebrew original has no Q-particle (20a), which again points to the use of the structure in non-translated, vernacular speech:

- (20) (Song of Salomon 3:3)
- a. *‘et še‘ahabá nafší re’item*
(Hebrew Bible)
 - b. ¿Al que amó la mi alma **si** lo vistes?
(BM-Santillana)
 - c. ¿Al que amó la mi ánima **si** vistes?
(BM-Arragel)
‘Have you seen him whom my soul loved?’

After the Middle Ages, direct interrogative *si* becomes the norm when translating Hebrew *ha* and *im* in the highly literal Bibles produced by Jewish exiles and their descendants after the expulsion of the Jews from Spain in 1492; otherwise, it largely disappears from Bible translation.¹⁹

5. Interrogative *si* in Medieval Romance

The use of *si* to introduce yes/no-Qs is not just a Spanish phenomenon. Despite Herman's claim that there is no continuity of direct interrogative *si* in the Romance languages, the structure is attested in Old French (Marchello-Nizia, 1985, pp. 220, 229–230) and Old Occitan (Jensen, 1990, p. 526, 1994, pp. 354–355). Historical grammars do not mention the existence of direct interrogative *si* in Galician-Portuguese, Catalan, or the medieval dialects of Italy, but again, we are dealing with a phenomenon that has passed virtually unnoticed among Romanists. Also, the interrogative particle *si* is not easily traceable in the available historical corpora because it is homophonous with other common words, such as the conditional conjunction or the affirmative adverb. In the absence of a detailed analysis we cannot rule out the existence of the structure in Catalan, Galician-Portuguese, or the Italian dialects.

In the case of French, it has been pointed out that direct interrogative *si* is rather infrequent (cf. Meyer-Lübke, 1900, § 519; Marchello-Nizia, 1985, pp. 222–221). Still, there is a good number of examples of *si/se* in direct questions (cf. (21a)–(21d)).

- (21) a. *Queus jorz, sire? si nel savez?*
 'What day, my lord? Don't you know it?' (Perceval, v. 6265, ca. 1180)
- b. *Se tu me sez dire noveles/Des chevaliers et des puceles?*
 'Could you tell me any news about those knights and those young ladies?' (Perceval, v. 279, ca. 1180)
- c. «*Dame*», *fet il, «s'il vos remanbre/Del nain qui ier vos correça/Quant vostre pucele bleça?*
 "'My lady", he said, "do you remember the dwarf who made you angry yesterday when he hurt your maiden?" (Erec, 1110, ca. 1176)
- d. *Se peussiez esclarcir ma merde,/maistre Jean?*
 'Could you give me a purge, master Jean?' (Pathelin, 666–669, ca. 1467)

19. However, it is possible to find a handful of examples in the 1569 *Biblia del Oso*.

Alt-Qs could be marked by *si* as well (22a)–(22b). In this case, the particle can appear in front of both conjoined interrogatives or only in front of one of them, either the first one, as in (22a), or the second one (22b).²⁰

- (22) a. *Fet l'empereres: «Se te vient/d'orgoeil ou de melancolie/que tu hez tant ma compegnie?»*
 'Is it out of pride or melancholy that you enjoy my company so little?'
 (*Roman de la Rose*, 645–647, ca. 1228)
- b. *avez vous donc robe achatee ou se vous l'avez empruntee*
 'Did you buy a dress then, or did you borrow it?'
 (*Méon* III 278.192, Jensen, 1990, p. 526)

As for Catalan, although historical grammars make no mention of this use there is at least one example in biblical translation in the 14th century Bible. In this particular passage the translator renders *si* in the Latin original with its closest Catalan equivalent, that is, *si*.

- (23) *Si tu lliures aquells en les mans del poble de Israel?*
 'Will you deliver them into the hand of Israel?'
 (Samuel 1: 14:37, Catalan Bible, 14th c.)

Likewise, we have not found information on this structure in historical grammars of Galician and Portuguese; a 13th century Galician-Portuguese song by Martin Codax, however, contains a few examples of direct interrogative *se*.

- (24) *Ondas do mar de Vigo, se vistes meu amigo? e ai Deus, se verrá cedo?*
 'Waves of the sea of Vigo, have you seen my boyfriend? Oh God, will he come soon?'
 (*Cantigas*, 513)

The use of direct interrogative *si* in Old French has been explained as the analogical extension of the conjunction *si* that introduces embedded interrogatives (Nyrop, 1925: V, § 31.5; 1930: VI, § 386; Jensen, 1990, p. 526). Another possible explanation, which is compatible with the analogy, is the continuation of Latin structures in which extension of *si* in indirect interrogatives was encouraged by the Hebrew calque that is present in Greek and Latin biblical translation.

20. Alt-Qs with *si* in front of the second member as in *Mon coeur court-il au change, ou si vous l'y poussez?* [Molière, *Les femmes savants*, v. 1187] (Nyrop, 1925, V, § 38) were frequent in French until the 17th century and are still used in written registers (Nyrop, 1930, VI, § 386).

6. The grammaticalisation of Q-particles

In the preceding sections we have argued that the historical origin of direct interrogative *si* in Romance is related to the extension of Latin conditional *si* to introduce embedded interrogatives and, eventually, become a Q-particle in direct questions. The diachronic process that we are, thus, proposing is:

Conditional > Indirect interrogative > Polar Q-marker

Figure 1. The diachronic development of the Q-particle *si*

This sort of development has resulted in the use of related markers for conditional, indirect interrogatives and yes/no-Qs in a number of languages (i.e. biblical Hebrew *im*, Greek *ei*, Latin *si*, Old Spanish, Old French and Old Occitan *si*). The same process can be observed in a few other languages: Yiddish yes/no-Qs are optionally introduced by *tsi* ‘if, whether; or’ (*tsi zogst du mir ersht itst?* ‘are you telling me now for the first time?’ (Jacobs et. al. 1994, p. 413)). Likewise, Swedish conditional sentences and embedded interrogatives are introduced by *om* ‘if/whether’ which can also introduce desiderative sentences expressing a wish; in the Swedish dialect spoken in Solf (Finland) *om* can even head yes/no-Qs (Andersson, 1994; Raukko & Östman, 1994). The use of *om* as a yes/no Q-particle has been reported as well for the Stavanger Norwegian Dialect (*Om du snakke stavanger- skt?* ‘Do you speak Stavanger dialect?’ (Lie, 1992, pp. 67–68)). A similar structure exists in German where *ob* ‘whether’ can introduce a direct interrogative: *Ob ich mal wegen meiner Galle frage?* ‘[what would you think], if I just ask about my gall bladder?’ (Buscha, 1976). Other strategies would involve a *wh*-item (namely, *what*) giving rise to polar Q in Marathi while a disjunctive marker gave rise to polar Q in Finnish.

The existence of Q-particles that derive from or are related to conditional structures is hardly surprising if we take into account that conditional sentences, desiderative structures, embedded interrogatives and yes/no-Qs are all utterances that express uncertainty and doubt, and such meanings are consistent with the function and pragmatics of questions. Leaving aside non-canonical questions, interrogative speech acts request information, and this is often formulated as a hypothesis for which confirmation is requested. Also, because questions involve making a request to the addressee, they are often formulated using hedging mechanisms to tone down the strength of the request (Bencini, 2003, p. 609). Conditional and desiderative structures are common devices used to mitigate requests; as a result they are liable to become sources for interrogative

structures.²¹ However, current surveys of the grammaticalisation of Q-particles do not consider conditional or desiderative as a common source for such markers. Heine & Kuteva (2002, pp. 216–217, pp. 226–227) and Bencini (2003) only list the negative marker NOT and the disjunctive operator OR as sources for polar interrogative markers, although, of course, this does not rule out other possible sources.²² In fact, contrary to what we observe in the history of Q-particle *si* –the evolution of a conditional marker into a Q-particle– numerous studies have identified processes that work in the opposite direction, that is, interrogatives that become a source for the grammaticalisation of conditional connectives (Haiman, 1978; ; Heine & Kuteva, 2002, p. 249; Heine & Siemund, 2001, pp. 1014–1015; Ohori, 2011, p. 643; Traugott, 1985, p. 249).

The historical process that we are suggesting, from indirect interrogative to yes/no-Q, entails a change from a complex structure (i.e. an embedded question) to a simple one (see Harris & Campbell, 1995, pp. 298–308 for the opposite case scenario). Crucially, less attention has been paid to this kind of change, namely subordination as the starting point for further changes that result in a main clause structure. Evans (2007) has identified a number of such structures in a wide array of languages and has coined the term “insubordination” to refer to this phenomenon defined as “the conventionalised main clause use of what, on prima facie grounds, appear to be formally subordinate clauses” (p. 367). He proposes that such structures emerge through a four-stage historical development, as shown in Table 3:

Table 3. Four-stage historical development for insubordination

A	B	C	D
<i>Subordination</i>	<i>Ellipsis</i>	<i>Conventionalised ellipsis</i>	<i>Reanalysis as main clause structure</i>
Subordinate structure	Ellipsis of main clause	Restriction of interpretation of elided material	Conventionalised main clause use of formally subordinate clause

As already pointed out, research is still rudimentary, and not much is known about the historical evolution and precise uses of Old Spanish direct interrogative *si*. But

21. Evans (2007, pp. 380, 390) gives examples of sentences that use equivalents of ‘if’ to introduce polite requests in French, English, Dutch and Basque.

22. Bencini (2003, p. 609) points at conditional or hypothetical markers as possible sources for Q-particles and reports that according to Nocentini (p. c.) many Meso-American languages seem to have Q-particles which are related to or derive from conditional markers.

at first sight, the stages in its development fit well in Evans's model: Q-particle *si* derives from a subordinate structure (the Latin indirect interrogative), which has become a main clause structure via ellipsis, and whose meaning has been restricted from expressing any polar interrogative in Medieval Spanish, to only desiderative and 'I wonder' type questions after the 1700s.

Moreover, insubordination phenomena seem to be common in the Romance languages as demonstrated by numerous examples in which complementisers become Q-particles. In many cases this can be traced to different types of complex structures such as cleft sentences, echo questions or subordinate clauses in which the matrix has been eliminated via ellipsis and the complementiser has evolved into a Q-particle. For instance, there are cases of complementisers that have become Q-particles in certain non-canonical interrogative structures in Modern Spanish, such as quotative *que*, *¿como qué?*, *¿conque?* or echo-question *si* (Escandell, 1999, pp. 3965–3972). Ledgeway (2012, pp. 177–178, p. 299) points out the grammaticalisation of complementisers as interrogative markers such as Ibero-Romance and Occitan *que* 'that, what', French *est-ce que* (erstwhile cleft: 'is this that...'), Florentine *o(cche)* 'or (what)' and central-southern Italian dialect *che/chì/ce* 'that/what'. To this we could add Portuguese Q-particle *é que* which more transparently incorporates the copula *é* in the same way French interrogative particle *est-ce-que*.

In brief, the developments that have facilitated the change from conditional *si* to a Q-marker (extension of conditional conjunctions to head embedded interrogatives and grammaticalisation of complementisers as Q-particles) are plausible changes and well attested in a number of languages. The historical processes that we are proposing here deserve further exploration within grammaticalisation theory, to see if they are counterexamples to the usual mechanisms in the formation of Q-particles, and/or if they are indeed relevant and productive.

7. Conclusions

In this article we have provided data and analysis for the historical origin of Spanish polar Q-particle *si*, a structure that has barely been described in the literature. We have shown that it originates in the Classical Latin conditional conjunction *si* which in Late Latin became a complementiser in embedded interrogatives and eventually a polar Q-particle. This evolution parallels similar developments in biblical Hebrew and biblical Greek, and is documented primarily in biblical translation and religious texts. However, its presence in non-religious texts and, more importantly, its continuation in Medieval Romance varieties show that direct interrogative *si* must have had some use in vernacular Latin. The historical

origin and evolution of Q-particle *si* constitutes an interesting case of evolution from conditional marker to complementiser for embedded questions and then to Q-particle in polar interrogatives, a process attested in other languages, but that has not been considered in accounts of the historical origin of Q-particles in grammaticalisation studies.

References

- Almeida, B. (2013). Fenómenos sintácticos raros y edición de textos: El caso de la Segunda Parte de la *General Estoria*. *Revista de Historia de la lengua Española*, 8, 3–28.
- Andersson, E. (1994). Swedish. In E. Köning & J. van der Auwera (Eds.), *The Germanic languages* (pp. 271–312). London: Routledge.
- Arnaiz, A. R. (1998). The main word order characteristics of Romance. In A. Siewierska (Ed.), *Constituent order in the languages of Europe* (pp. 47–73). Berlin: Mouton de Gruyter. doi:10.1515/9783110812206.47
- Bencini, G. (2003). Towards a diachronic typology of yes/no question constructions with particles. *Proceedings of the thirty-ninth annual meeting of the Chicago Linguistic Society* (pp. 604–621). Chicago, IL: Chicago Linguistic Society.
- Bertocchi, A., & Maraldi, M. (2011). Conditionals and concessives. In P. Baldi & P. Cuzzolin (Eds.), *New perspectives on historical Latin syntax 4. Complex sentences, grammaticalization, typology* (pp. 661–864). Berlin: Mouton de Gruyter. doi:10.1515/9783110253412.93
- Blaise, A. (1954). *Dictionnaire latin-français des auteurs chrétiens*. Turnhout: Brepols.
- Blaise, A. (1955). *Manuel du latin chrétien*. Strasbourg: Le latin chrétien.
- Blass, F., & Debrunner, A. (1982). *Grammatica del greco del Nuovo Testamento*. F. Rehkopf (Ed.). Brescia: Paideia.
- Bodelot, C. (1987). *L'interrogation indirecte en Latin: Syntaxe – valeur illocutoire – formes*. Paris: Société pour l'information grammaticale.
- Bodelot, C. (1990). *Termes introducteurs et modes dans l'interrogation indirecte en latin de Plaute a Juvenal*. Avignon: Aubanel.
- Bouzouita, M. (2008). *The diachronic development of Spanish clitic placement*. (Unpublished doctoral dissertation) King's College, University of London, London, United Kingdom.
- Bouzouita, M. (2011). Future constructions in Medieval Spanish: Mesoclisix uncovered. In R. Kempson, E. Gregoromichelaki & C. Howes (Eds.), *The dynamics of lexical interfaces* (pp. 91–132). Stanford, CA: CSLI.
- Brown, H. P., Joseph, B. D., & Wallace, R. E. (2009). Questions and answers. In P. Baldi & P. Cuzzolin (Eds.), *New perspectives on historical Latin syntax 1. Syntax of the sentence* (pp. 489–530). Berlin: Mouton de Gruyter.
- Buridant, C. (2000). *Grammaire nouvelle de l'ancien français*. Strasbourg: Sedes.
- Buscha, A. (1976). Isolierte Nebensätze im dialogischen Text. *Deutsch als Fremdsprache*, 13, 274–279.
- Clines, D. J. A. (1993–2012). *The dictionary of Classical Hebrew*. Sheffield: Sheffield Academic Press.
- Dryer, M. S. (2011). Polar Questions. In M. Haspelmath, M. S. Dryer, D. Gil & B. Comrie (Eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library. Retrieved from <<http://wals.info/feature/116>>

- Elsig, M. (2009). *Grammatical variation across space and time: The French interrogative system*. Amsterdam: John Benjamins. doi:10.1075/silv.3
- Enrique-Arias, A. (2006). Sobre el parentesco entre la *Biblia de Alba* y la *Biblia* de la Real Academia de la Historia ms. 87. *Romance Philology*, 59, 21–43. doi:10.1484/J.RPH.2.303252
- Enrique-Arias, A. (2008). Apuntes para una caracterización de la morfosintaxis de los textos bíblicos medievales en castellano. In J. Kabatek (Ed.), *Sintaxis histórica del español y cambio lingüístico: Nuevas perspectivas desde las tradiciones discursivas* (pp. 109–125). Frankfurt & Madrid: Vervuert & Iberoamericana.
- Enrique-Arias, A., & Burguera, J. (2010). Variación y cambio en la formalización de la interrogación retórica en la historia del español. Presentation given at the *XXVI Congreso Internacional de Lingüística y Filología Románicas*, Universidad de Valencia, 6–11 September 2010.
- Ernout, A., & Thomas, F. (1953). *Syntaxe latine*. Paris: Klincksieck.
- Escandell Vidal, M. V. (1999). Los enunciados interrogativos. Aspectos semánticos y pragmáticos. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (pp. 3929–3991). Madrid: Espasa-Calpe.
- Evans, N. (2007). Insubordination and its uses. In I. Nikolaeva (Ed.), *Finiteness: Theoretical and empirical foundations* (pp. 366–431). Oxford: Oxford University Press.
- Fava, E. (1995). Il tipo interrogativo. In L. Renzi, G. Salvi & A. Cardinaletti (Eds.), *Grande grammatica italiana di consultazione. Volume III. Tipi di frase, deissi, formazione delle parole* (pp. 70–126). Bologna: Il Mulino.
- García de Diego, V. (1951). *Gramática histórica española*. Madrid: Gredos.
- Gordis, R. (1933). A rhetorical use of interrogative sentences in biblical Hebrew. *The American Journal of Semitic Languages and Literatures*, 49(3), 212–217. doi:10.1086/370382
- Haiman, J. (1978). Conditionals are topics. *Language*, 54, 512–540. doi:10.1353/lan.1978.0009
- Harris, A. C., & Campbell, L. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620553
- Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511613463
- Herman, J. (1996). A propos du *si* interrogatif: Évolutions achevées et évolutions bloquées. In A. Bammesberger & F. Heberlein (Eds.), *Akten des VIII. Internationalen Kolloquiums zur lateinischen Linguistik* (pp. 296–307). Heidelberg: Winter.
- Hofmann, J. B., & Szantyr, A. (1965). *Lateinische Syntax und Stylistik*. München: C. H. Beck'sche Verlagsbuchhandlung.
- Jacobs, N. G., Prince, E. F., & van der Auwera, J. (1994). Yiddish. In E. Köning & J. van der Auwera (Eds.), *The Germanic languages* (pp. 388–419). London: Routledge.
- Jensen, F. (1990). *Old French and comparative Gallo-Romance syntax*. Tübingen: Max Niemeyer. doi:10.1515/9783110938166
- Jensen, F. (1994). *Syntaxe de l'ancien occitan*. Tübingen: Max Niemeyer. doi:10.1515/9783110936148
- Köning, E., & Siemund, P. (2007). Speech acts distinctions in grammar. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol 1; pp. 276–324). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511619427.005
- Ledgeway, A. (2012). *From Latin to Romance. Morphosyntactic typology and change*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199584376.001.0001
- Lie, S. (1992). Ka du sei?. *Maal og Minne*, 1, 62–77.
- Löfsted, E. (1911). *Philologischer kommentar zur Peregrinatio Aetheriae*. Uppsala: Almqvist & Wiksell.
- Maiden, M., & Robustelli, C. (2007). *A reference grammar of Modern Italian*. London: Hodder Education.

- Marchello-Nizia, Ch. (1985). *Dire le vrai: L'adverbe "si" en français médiéval. Essai de linguistique historique*. Genève: Librairie Droz.
- Mira Mateus, M. H., Brito, A. M., Duarte, I., & Hub Faria, I. (2006). *Gramática da língua portuguesa*. Lisboa: Caminho.
- Nyrop, K. (1930). *Grammaire historique de la langue française*. Copenhagen: Gyldendalske Boghandel.
- Obenauer, H.-G. (2004). Nonstandard *wh*-questions and alternative checkers in Pagotto. In H. Lohnstein & S. Trissler (Eds.), *Syntax and semantics of the Left Periphery, Interface explorations* 9 (pp. 343–384). Berlin: Mouton de Gruyter. doi:10.1515/9783110912111.343
- Ohuri, T. (2011). The grammaticalization of subordination. In H. Narrog & B. Heine (Eds.), *The Oxford handbook of grammaticalization* (pp. 636–645). Oxford: Oxford University Press.
- Payrató, L. (2002). Lenunciació i la modalitat oracional. In J. Solà, M. -R. Lloret, J. Mascaró & M. Pérez Saldanya (Eds.), *Gramàtica del català contemporani* (pp. 1149–1220). Barcelona, Empúries.
- Pérez Saldanya, M., & Hualde, J. I. (2018). Recurrent processes in the evolution of concessive subordinators in Spanish and Catalan. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 223–248). Amsterdam: John Benjamins. (in this volume) doi:10.1075/ihll.16.10per
- Posner, R. (1995). Contact, social variants, parameter setting, and pragmatic function: An example from the history of French syntax. In J. Fisiak (Ed.), *Linguistic change under contact conditions* (pp. 217–236). Berlin: Mouton de Gruyter. doi:10.1515/9783110885170.217
- Pueyo Mena, F. J., & Enrique-Arias, A. (2013). Los romanceamientos castellanos de la Biblia Hebrea compuestos en la Edad Media: Manuscritos y traducciones. *Sefarad*, 73(1), 165–224. doi:10.3989/sefarad.013.006
- Raukko, J., & Östman, J.-A. (1994). *Pragmaattinen näkökulma itämeren kielialueeseen*. Helsinki: University of Helsinki.
- Real Academia Española/Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.
- Sachs, G. (1948–1949). Fragmento de un estudio sobre la *Biblia medieval romanceada*. *Romance Philology*, 2, 217–228.
- Sanchis Calvo, M. C. (1991). *El lenguaje de la Fazienda de Ultramar*. Madrid: Real Academia Española.
- Sephiha, H. V. (1989). Ladinismes dans *La Fazienda de Ultramar* du XII^{ème} siècle. In D. Tollet (Ed.), *Politique et religion dans le judaïsme ancien et médiéval, interventions au colloque des 8 et 9 décembre 1987* (pp. 319–325). Paris: Desclée.
- Siemund, P. (2001). Interrogative constructions. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (Eds.), *Language typology and language universals: An international handbook* (Vol. 2; pp. 1010–1028). Berlin: Walter de Gruyter.
- Traugott, E. C. (1985). Conditional markers. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 289–307). Amsterdam: John Benjamin. doi:10.1075/tsl.6.14clo
- Väänänen, V. (1981). *Introduction au latin vulgaire, troisième édition revue et augmentée*. Paris: Klincksieck.
- Wheeler, M. W., Yates, A., & Dols, N. (1999). *Catalan: A comprehensive grammar*. London: Routledge.
- Wright, R. (2018). Studying Ibero-Romance before 1200. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 325–335). Amsterdam: John Benjamins. (in this volume) doi:10.1075/ihll.16.14wri

Corpus

- Amadís* = Garcí Rodríguez de Montalvo, *Amadís de Gaula*, J. Manuel Cacho Bleuca (Ed.). Madrid: Cátedra, 1991.
- Atalaya* = Alfonso Martínez de Toledo, *Atalaya de las crónicas*, Ms. Egerton 287, J. B. Larkin (Ed.), ADMYTE-II, 1999.
- Bienandanzas* = Lope García de Salazar, *Bienandanzas e fortunas*, A. M. Marín Sánchez (Ed.). Retrieved from <parnaseo.uv.es/Lemir/textos/bienandanzas/Menu.htm>.
- BM* = Enrique-Arias, A. (2008–2014). *Biblia Medieval corpus*. Retrieved from <http://www.bibliamedieval.es>
- Buen amor* = Juan Ruiz, *Libro de buen amor*, A. Bleuca (Ed.). Madrid: Cátedra, 1992.
- Cantigas* = 500 *Cantigas d'amigo*. R. Cohen (Ed.). Porto: Campo das letras, 2003.
- Catalan Bible* = *Biblia del segle XIV: Primer i segon llibre dels Reis*, J. Bruguera Talled, P. Casanelas & N. Calafell Sala (Eds.). Barcelona: Associació Bíblica de Catalunya/Publicacions de l'Abadia de Montserrat, 2011.
- Celestina* = Fernando de Rojas, *La Celestina*, D. S. Severin (Ed.). Madrid: Cátedra, 1987.
- Celos* = Miguel de Cervantes, *El gallardo español; La casa de los celos*, F. Sevilla Arroyo, & A. Rey Hazas (Eds.). Madrid: Alianza, 1997.
- Cid* = *Cantar de mio Cid. Texto, gramática y vocabulario*, Vol. 3: *Texto*, R. Ménezdez Pidal (Ed.). Madrid: Espasa-Calpe, 1980.
- Corbacho* = Alfonso Martínez de Toledo, *El Corbacho*, E. W. Naylor (Ed.), ADMYTE-II (ms. h-III-10), 1999.
- Erec* = Chrétien de Troyes, *Érec et Énide*. Paris: Flammarion, 2009.
- GEI* = Alfonso X, *General estoria. Primera parte*, P. Sánchez-Prieto Borja (Ed.). Madrid: Biblioteca Castro, 2009.
- Greek Septuaginta* = *Septuaginta*, A. Rahlfs, & R. Hanhart (Eds.). Stuttgart: Deutsche Bibelgesellschaft, 2006.
- Hebrew Bible* = *Biblia Hebraica Stuttgartensia*, K. Elliger, & W. Rudolph (Eds.). Stuttgart: Deutsche Bibelgesellschaft, 1997.
- Isla* = José Francisco de Isla, *Historia del famoso predicador Fray Gerundio de Campazas, alias Zotes*, J. Jurado (Ed.). Madrid: Gredos, 1992.
- Latin Vulgate* = *Biblia Sacra Vulgata*, R. Weber, & R. Gryson (Eds.). Stuttgart: Deutsche Bibelgesellschaft, 1994.
- Lazarillo* = *Lazarillo de Tormes*, F. Rico (Ed.). Madrid: Real Academia Española, 2011.
- Liv.* = *Livy, with an English translation in fourteen volumes. VIII, books XXVIII-XXX*, F. Gardner Moore (Ed.). London/Cambridge: Harvard University Press, 1962.
- Pathelin* = *Maistre Pierre Pathelin; farce du XVe siècle*, R. T. Holbrook (Ed.). Paris: H. Champion, 1924.
- Perceval* = Chrétien de Troyes, *Perceval ou le Conte du graal*. Paris: Flammarion, 1997.
- Peribáñez* = Félix Lope de Vega, *Peribáñez y el comendador de Ocaña*, D. McGrady (Ed.). Barcelona: Crítica, 1997.
- Pl. Amph.* = *Plautus, with an English translation by Paul Nixon in five volumes. I. Amphitryon, The comedy of asses, the pot of gold, the two bacchises, the captives*. London/Cambridge: Harvard University Press, 1966.
- Pl. Trin.* = *Plautus, with an english translation by Paul Nixon in five volumes. V. Stichus. Three Bob Day, Truculentus, The tale of a travelling bag, fragments*. London/Cambridge: Harvard University Press, 1955.

- Plácidas* = *El caballero Plácidas* (Ms. Esc. H-I-13), R. M. Walker (Ed.). Exeter: University of Exeter, 1982.
- Quijote* = Miguel de Cervantes, *Don Quijote de la Mancha*, F. Rico (Ed.). Madrid: Instituto Cervantes/Crítica, 1998.
- Roman de la Rose* = Jean Renart, *Le Roman de la rose ou de Guillaume de Dole*, F. Lecoy (Ed.). Paris: Champion, 1979.
- Regenta* = Leopoldo Alas "Clarín", *La Regenta*, G. Sobejano (Ed.). Madrid: Castalia, 1981.
- Ter., Eu. = Terence. I. *The woman of Andros, the self-tormentor, the eunuch*, J. Barsby (Ed.). London/Cambridge: Harvard University Press, 2001.
- Tert. Apol. = Tertulian. *Apology. De spectaculis*, T. R. Glover, & G. H. Rendall (Eds.). London/Cambridge: Harvard University Press, 1984.
- Zalacaín* = Pío Baroja, *Zalacaín el aventurero*. Madrid: Caro Raggio, 1975.

Realmente, verdaderamente and ciertamente

On the relation between epistemic modality and intensifying adverbs

Mónica González Manzano

Universitat de Barcelona

This paper deals with the historical evolution of three epistemic adverbs in Spanish, *verdaderamente*, *realmente* and *ciertamente*. First, we describe how these adverbs grammaticalise – or pragmaticalise – as discourse markers and, therefore, widen their structural scope. Secondly, we show how these words experience a subsequent functionalisation, allowing them to function once again as adverbs, but now as adjective modifiers and not as verbal adverbs. Finally, we conclude that the fact that *verdaderamente*, *ciertamente* and *realmente* preserve all these different meanings at the same time is not out of line with many other grammaticalisation processes, a feature defined in the bibliography as “layering” (Hopper, 1991).

Keywords: Spanish; grammaticalisation; epistemicity; adverbs; discourse markers; intensifying

1. Introduction

Before describing the historical evolution of the three epistemic adverbs, namely *verdaderamente*, *realmente* and *ciertamente* in Spanish, we should mention that the behaviour of *-mente* adverbs has been widely discussed in the bibliography as they show some specific morphosyntactic features.¹ It has always been argued that

1. For further details about the development of *-mente* adverbs in Spanish as well as in other Romance languages, see Karlsson (1981). For further information about the grammatical status and morphosyntactic behaviour of *-mente* adverbs in Spanish, see Rodríguez Ramalle (2003), Egea (1979), Varela (1990) and Kovacci (1999).

these features are caused by the etymological origin of these adverbs, given that they developed in vulgar Latin as a construction formed by the combination of an adjective and the noun *MENS* in the ablative case, *MENTE*, with a meaning similar to ‘with this state of mind, having this specific predisposition.’ This structure completely replaced classic Latin adverbial derivative suffixes because the previous system was highly polyfunctional and confusing for the speaker. Eventually, this periphrastic structure was reanalysed as a synthetic one in the Romance languages where it is still highly productive.

Given this periphrastic origin, *-mente* adverbs in Spanish show a particular syntax and morphology. So, in spite of being formally a single word, they preserve some of the features of their old pluriverbal status in Latin, as indicated: “[el] mantenimiento de la acentuación prosódica de los dos constituyentes” (‘the maintaining of the prosodic stress of the two constituents’), “la posibilidad de omitir en una coordinación el primer *mente*” (‘the possibility to omit the first *-mente* in coordinated [adverbs of this type]’) and “la presencia de la marca femenina singular del adjetivo, concordante con el nombre *mente*” (‘the presence of the feminine singular ending that agrees with the noun *mente*’) (Rodríguez Ramalle, 2003, p. 19).² Besides these syntactic particularities, they show the singularity of modifying the whole meaning of the sentence, which will be considered as an essential feature in order to understand their subsequent pragmaticalisation as discourse markers.

Another important theoretical issue concerns the grammatical status of *-mente* adverbs once they grammaticalise with more pragmatic meanings, as well as whether they can be considered as discourse markers or not. In some of the first discourse marker studies, especially those developed in the framework of Relevance Theory,³ epistemic and deontic adverbs were not considered as such, as they preserve a conceptual, lexical meaning, even when showing a pragmatic function, whereas discourse markers were defined as exhibiting only procedural

2. That is to say, we can verify the periphrastic origin of *-mente* adverbs through the following facts:

1. When pronounced, the stress still falls on two different phonemes, although we can only see one graphic accent (*difícilménte* ‘with difficulty’, for example).
 2. They preserve an *-a* before the *-mente* suffix, which is the morphological mark for the feminine form of the adjective. This shows agreement with the feminine noun *mente*.
 3. When two *-mente* adverbs are coordinated, the first one can be omitted, as, for example, in *simple y llanamente* ‘simply and plainly’ (instead of *simplemente y llanamente*).
3. For a Relevance Theory approach to discourse markers, see Blakemore (1987, 1992, 2002) and Wilson and Sperber (1993).

meaning. However, epistemic adverbs have recently been considered as pragmatic markers,⁴ given that they imply highly subjective meanings, such as the speaker's certainty and reinforcement, for example.

In this paper we have considered the epistemic adverbs *verdaderamente*, *ciertamente* and *realmente* as discourse markers when they acquire the possibility of functioning in a peripheral position with a pragmatic meaning, although the original modal meaning is never completely erased.

This paper is organised in the following way: in Section 2 we describe the data as well as the results from electronic corpora we have worked with. In Section 3, we show the syntactic behaviour and semantic meaning of *verdaderamente*, *ciertamente* and *realmente* when they function as adverbs of manner, that is to say, when they appear in a postverbal position with a function that has been defined as a verbal adverb. In Section 4, we see how these verbal adverbs can easily occupy a preverbal position, with scope over the whole statement. We will define this function as the sentence adverb. Section 5 shows how these sentence adverbs can move towards a more peripheral position consolidating a more pragmatic meaning. In Section 6 we see how these epistemic adverbs developed another intra-sentential function as adjective modifiers. Finally, in the conclusions, we summarise the most relevant changes described.

2. Methodology

The present paper deals with three epistemic adverbs in Spanish, *verdaderamente*, *realmente* and *ciertamente*. We have chosen these forms because they share three main features: (a) they express epistemic modality, which constitutes a semantic field that shows a strong tendency towards pragmatics, (b) they share the suffix *-mente* and, therefore, some special morphological and syntactic features, (c) not only do they undergo grammaticalisation as discourse markers, but also, ultimately, a functionalisation process as intensifying adverbs.

These three adverbs will help us show how epistemic modality is a major source for the development of pragmatic (speaker-oriented and addressee-oriented) functions, as well as how epistemicity relates to intensification.

In order to analyse the historical trajectory of *verdaderamente*, *realmente* and *ciertamente*, we have extracted data from two electronic corpora, CORDE and CREA, both designed by the Real Academia Española. The shortcomings and

4. Martín Zorraquino (2010), for example, includes both epistemic and deontic adverbs in her most recent classification of discourse markers.

limitations of CORDE and CREA are well attested in the bibliography,⁵ but, as Estellés (2009, p. 51) affirms, in order to understand grammaticalisation processes, which can last several centuries, it is necessary to consult large corpora and CORDE offers an extensive number of documents which cover the history of Spanish from the 13th to the 20th c. It is also worth mentioning that many different genres (narrative, legal texts, poetry, etc.) are also included in this corpus, thus, offering a fairly complete overview of the history of these adverbs in the Spanish language. We have also included some examples from CREA, when it was necessary to look at more recent data since this corpus covers all kinds of registers from 1980 onwards. We chose a large number of random examples from these corpora to analyse:⁶

Table 1. Number of examples extracted from CORDE

Verdaderamente	4.794 examples
Realmente	2.790 examples
Ciertamente	2.542 examples

Table 2. Number of examples extracted from CREA

Verdaderamente	922 examples
Realmente	826 examples
Ciertamente	826 examples

3. *Verdaderamente*, *realmente* and *ciertamente* as verbal adverbs

When *verdaderamente*, *realmente* and *ciertamente* appear immediately to the left of the verb, they usually express semantic notions related to modality. In

5. For a critical review of the limitations that arise from the fact that CORDE has not been lemmatised nor morphologically tagged, see Lleal (2010). For a comparison between two big electronic corpora in Spanish, CORDE and *Corpus del español*, see Davies (2008) and Estellés (2009, p. 51 and ff.). The latter also offers a justification for the use of digital corpora in grammaticalisation studies.

6. One of the issues that requires further research is the impact of diatopic variation in the different uses of the adverbs analysed. We are aware that more attention has to be paid to data from different Spanish-speaking countries, but for the present paper, we prefer to focus on historical variation only. For this reason, we have chosen only examples from CORDE and CREA that are from Spain.

this syntactic position, they strictly modify the verb, that is, they take scope over the VP.

This function has usually been defined as a verbal adverb or a predicate adverb (see Dik et al., 1990; Dik, 1997; Ramat & Ricca, 1998; Traugott, 1995). In the framework of Functional Grammar by Simon Dik and Kees Hengeveld (see Dik et al., 1990; Hengeveld, 1989), the verbal adverb – or “predicate satellite”, in Dik’s terminology – is considered a satellite that functions in the representational level. As Ramat and Ricca (1998, p. 190) specify, verbal adverbs can be related to “the event (or state of affairs) denoted by the sentence”. That is to say, they can be defined as σ_1 satellites (see also Dik et al., 1990, p. 29).

For the purpose of this paper, we have been able to document examples of verbal adverbs *verdaderamente* and *ciertamente* since the 13th–14th c. Therefore, we considered this verbal modifying function as the core meaning of these adverbs. However, we must note that, whereas *verdaderamente* and *ciertamente* seem to be consolidated as adverbs from the 14th c., we have not been able to document *realmente* prior to the 15th c. This difference is more easily understood when observing the etymological origins of each word. *Verdaderamente* and *ciertamente* were formed by adding *-mente* to the feminine forms of the adjectives *verdadero* and *cierto*. Both words were introduced in Spanish before the 13th c. and were, at that time, considered popular terms.⁷ By contrast, the adjective *real* was introduced in Spanish during the 15th c., as a Latin term, only used in formal registers. That explains the appearance of *realmente* from the 15th c. onwards, in a more restricted use, if we compare it to *verdaderamente* and *ciertamente*. *Realmente* was, in fact, a learned word that was not widely used until the 17th–18th c.⁸

When functioning as verbal adverbs, we can see that they express a modal notion, although a very specific kind of modality:

7. Note that in the *Diccionario crítico y etimológico castellano e hispánico* (DCECH, Corominas & Pascual ([1980] 1991) *verdadero* is first attested in 1029, and *cierto* during the second half of the 10th c. Unfortunately, such old documents are not included in CORDE. If we search *cierto* and *verdadero* in this corpus, we will find a considerable number of examples during the second half of the 13th c., which provides evidence of the widespread use of these terms in medieval Spanish.

8. The information provided by DCECH confirms our intuition about the more learned use of *realmente*. In this dictionary, the adjective *real* (with the meaning of ‘having an effective existence’) is first attested in 1607, though *real*, with the meaning of ‘belonging to the king’ is first documented in 1188. We have been able to find examples of *real* with the former meaning before the date proposed by DCECH during the 15th c.

- (1) *gran cargo toma sobre sí para fazer lo que deve sin reprehensión, guardando su ley verdaderamente*
 ‘He takes great responsibility, in order to do what he should without being reprehended, truly keeping his law’
 (Anónimo, *Libro del cavallero Zifar*, 1300–1305; CORDE)
- (2) *de los quales dichos diez mill maravedís el dicho Lope de Alemaña se otorgó por entrego e bien pago, por quanto los resçebió realmente e con effecto en presençia de mí*
 ‘The above-mentioned Lope de Alemaña declared that he had been given and had been paid ten thousand *maravedís*, and that he really received them with effect and in my presence’
 (Anónimo, *Documentación medieval de la iglesia catedral de León*, 1419–1426; CORDE)

We can see in (1) that *verdaderamente* cannot be paraphrased exactly as ‘in a real, accurate and reliable manner’, as other *-mente* adverbs can be (for example *rápídamente* means ‘in a fast, rapid manner’), because this adverb does not express a physical manner. In other words, *verdaderamente*, more than merely stating a fact, affirms the certainty of the speaker about the truth of his words. Also, *realmente*, in (2) can be read as expressing how the verbal action has taken place. However, given the abstract notions that *real* and *verdadero* imply, *realmente* and *verdaderamente* do not work as adverbs of manner in the same way as *rápídamente* or *generosamente*.⁹ This fact will help us understand their subsequent function as sentence adverbs.

We observe that the number of verbs that tend to co-occur with these adverbs is relatively limited. In the case of *verdaderamente* and *ciertamente*, the co-occurrence with epistemic and speaking verbs is quite relevant,¹⁰ especially between the 13th and the 16th c., which is, as we will see in Section 5, when the discourse marker

9. That is to say, as verbal adverbs, they show epistemic modality, which, according to Traugott (1989, p. 32), “has to do with knowledge and belief about possibilities, probabilities, and so forth”. For more definitions of epistemic modality, see also Lyons (1977, p. 793), or Palmer (1986). This means that the adverbs analysed are semantically related to the expression of the speaker’s evaluation of the certainty of what they are going to say, and not to the expression of the manner of doing something.

10. In the case of *verdaderamente*, it co-occurs with an epistemic or a speaking verb in 20% of the examples documented between the 13th and the 16th c. In the case of *ciertamente*, this co-occurrence was found in 33% of the examples documented between the 13th and the 14th c., and in 26% between the 15th and the 16th c. In both cases, this high frequency clearly decreases during the subsequent period analysed: between the 17th and 18th c., *verdaderamente* co-occurs with epistemic and speaking verbs in 7% of the examples, and *ciertamente* in 6%, approximately.

function will arise and consolidate (see Pountain, in this volume, for more information on other discourse markers in the 16th c.). In the case of *realmente*, this verbal typology does not show a relevant frequency of co-occurrence. As a verbal adverb, *realmente* tends to appear in contexts like (3), which belong to legal texts, where the author wants to highlight that a transaction namely, a sale, a payment or a sentence, has been fully accomplished, in a physical, material way. That is also why it often appears in legal formulas, followed by adverbial phrases such as *de fecho* ‘in fact’:

- (3) *de los quales dichos dozientos moravedís me otorgo por bien contento e pagado a toda mi voluntad por quanto los recibí realmente e de fecho e pasaron a mi libre poderío*

‘I consider myself satisfied and I declare that I have been paid two hundred *maravedís*, and therefore I really received them and in fact they came into my free possession’

(Anónimo, *Carta por la que Pedro Guerrero vende a Teresa Rodríguez*, 1446; CORDE)

Finally, we have to remark that the three adverbs analysed appear to be functioning as verbal adverbs during the first periods considered. However, as we can see in Tables 3 to 5, this tendency has experienced a decrease from the 15th c. onwards.

In Table 3, we compare the use of *verdaderamente* with verbal structural scope to its use with non-verbal structural scope.¹¹ The results confirm that during the first period analysed (13th and 14th c.), this form tended to appear inside the sentence, taking scope over the VP. The 15th c. marks a point of change towards non-verbal structural scope:

Table 3. The scope of *verdaderamente*

	13th – 14th c.	15th – 16th c.	17th – 18th c.	19th – 20th c.
Verbal structural scope	68% (99/146)	46% (848/1853)	39% (333/842)	75% (1476/1955)
Non-verbal structural scope	32% (47/146)	54% (1005/1853)	61% (509/842)	25% (479/1955)

11. We have considered that these adverbs show a verbal structural scope when they function either as verbal adverbs, as in examples (1) to (3), or as adjective modifiers, as in examples like *las cosas verdaderamente/ciertamente/realmente amadas* (‘the really/truly loved things’), i.e. when they exhibit intra-sentential scope. By contrast, they show non-verbal structural scope when they belong to the extra-propositional level, i.e. when they function either as sentence adverbs or as discourse markers.

As we can see, contexts where *verdaderamente* has non-verbal structural scope increased from 32% of the examples documented during the 13th–14th c. period to 46% during the 15th–16th c. However, this tendency was reversed from the 18th c. onwards. This ultimate change is related to the consolidation of a new adverbial function, as an adjective modifier, as we will see in Section 6.

In Table 4, we can observe that *ciertamente* functions predominantly with verbal structural scope during the 13th–14th c. and begins to appear mostly with non-verbal structural scope from the 15th–16th c. onwards, showing a similar – if less dramatic – pattern to that of *verdaderamente*:

Table 4. The scope of *ciertamente*

	13th – 14th c.	15th – 16th c.	17th – 18th c.	19th – 20th c.
Verbal structural scope	58% (76/ 131)	33% (192/583)	29% (120/418)	32% (441/1374)
Non-verbal structural scope	42% (55/131)	67% (391/583)	71% (298/418)	68% (933/1374)

In the case of *realmente*, we can see in Table 5 that it shows the same historical evolution as *verdaderamente* and *ciertamente*, but with a different chronology. *Realmente* appears basically as an adverb with verbal structural scope (in 70% of the examples documented) until the 15th–16th c. The point of inflection, that is, the change towards the non-verbal structural scope, does not appear until the 17th c., as we can see in Table 5:

Table 5. The scope of *realmente*

	13th – 14th c.	15th – 16th c.	17th – 18th c.	19th – 20th c.
Verbal structural scope	67% (2/3)	70% (311/442)	20% (111/556)	47% (824/1765)
Non-verbal structural scope	33% (1/3)	30% (131/442)	80% (445/556)	72% (941/1765)

We can conclude then that *realmente* gains the use of a non-verbal structural scope more significantly in the 17th and 18th c. We have to relate this chronology, once more, with the introduction of *realmente* directly from Latin as a learned word into Spanish during the 15th c., much later than *verdaderamente* and *ciertamente*.

4. *Verdaderamente, realmente* and *ciertamente* as sentence adverbs

In Section 3, we defined the meaning of verbal adverbs *verdaderamente*, *realmente* and *ciertamente* as abstract since they express epistemic modality. This explains why in contexts like (4) these adverbs can be interpreted either as verbal adverbs or as adverbs with wider structural scope:

- (4) *El amor e la virtud dizen los sabios gentiles que tornan las cosas a ti. E si tal fuese el tuyo, naturalmente me convertiría yo a él. ¿E cómo quieres, Felipo, que yo ame a quien verdaderamente me avorreçe?*
 ‘Wise men say that love and virtue will give you back the things you have missed. And if your love was like that, I would naturally turn myself to it. And, how do you expect, Felipo, that I love the one who *really* detests me?’
 (Lope García de Salazar, *Istoria de las bienandanzas e fortunas*, 1471–1476; CORDE)
- (5) *En sustançia. Ase de entender en vna sustançia, porque, segunt los doctores theólogos, en esençia sustançia, o en Dios e natura, realmente son vna cosa, avnque los respectos sean diferentes*
 ‘According to the theologians, essence and substance, or God and nature, really/in fact are the same thing, although the proportions are different’
 (Gómez Manrique, *Poesías* [Cancionero de Gómez Manrique], ca. 1445–1480; CORDE)

In (4) *verdaderamente* appears in a preverbal position. However, it can still be understood as a verbal adverb, expressing the authentic manner of detesting someone, as well as a sentence adverb. In that case, *verdaderamente* contrasts two different situations, one, unreal and untrue where Felipo does not hate the female character, and another one, true and authentic where he really rejects her. Furthermore, in (5) we can read *realmente* as a verbal adverb that takes scope over *son* and expresses the authentic, reliable manner by which God and nature are the same thing, or it can also be interpreted as expressing reinforcement. Thus, *realmente* emphasizes the fact that God and nature, although having different proportions, are one and the same thing.¹²

Given this ambiguity, these adverbs consolidate a sentence adverb function. As we have seen, the syntactic position is a relevant factor that distinguishes between an adverb of manner reading and a sentence adverb one. Jackendoff (1972, p. 56) pointed out this different interpretation and proposed that for the adverbs “which have different readings depending on position, such as *cleverly*

12. Note that the presence of a verb like *ser* (‘to be’), which is semantically emptier than other verbs, facilitates the interpretation of *realmente* as a sentence adverb.

and *clumsily*, one reading is usually a manner adverb. In the other reading, some are speaker-oriented.” In the case of *verdaderamente*, *realmente* and *ciertamente*, we have seen that in a postverbal position, they can be read as manner adverbs – epistemic adverbs, whereas in a preverbal position they take scope over the whole sentence and express the speaker’s degree of certainty.

Although adverbs like *verdaderamente*, *ciertamente* and *realmente* have been described extensively in the bibliography as showing a relatively high degree of syntactic mobility, and although it is possible to find examples of preverbal adverbs with a verbal adverb function, we should highlight that there is a robust correspondence between a postverbal position and an intra-sentential scope, and that it is not very frequent that a verbal adverb appears in a preverbal position.¹³

By contrast, when *verdaderamente*, *ciertamente* and *realmente* function in an extra-propositional level, as sentence adverbs or as discourse markers, the preference for a preverbal position is quite strong, particularly in the case of *verdaderamente* and *realmente*. The former maintains a preference for a preverbal position in approximately 85% of the examples from the 13th to the 20th c., and the latter appears in almost 100% of the examples in a preverbal position until the 18th c. During the 19th and 20th c., the discourse marker *realmente* appears in a postverbal position in 22% of the examples.

However, in the case of *ciertamente* the preverbal position is the one preferred until the 17th–18th c., after which the postverbal position expands to more contexts, and finally the discourse marker *ciertamente* appears in a postverbal position in 54% of the examples documented during the 19th and 20th c.¹⁴ This last change in the syntactic position of *ciertamente* can be seen as an extension of the syntactic abilities of the adverb *ciertamente* once it has grammaticalised as a discourse marker.

In sum, although the syntactic position cannot be used as the only criterion to rule out an interpretation as a verbal adverb or as a sentence adverb, empirical evidence attests that more peripheral, extra-propositional elements tend to appear in a preverbal position.

The sentence adverbs, *verdaderamente*, *ciertamente* and *realmente* can be considered as disjunct adverbs, as is the case in Greenbaum (1969, p. 24 and ff.). In

13. None of the adverbs analysed in a preverbal position, when behaving as verbal adverbs, show a frequency of use higher than 10% through history.

14. *Ciertamente* and *verdaderamente* share many features and a similar chronology in their changes. However, *ciertamente* shows some particular changes, such as a complete reversal in the position of the discourse marker. For further information about the peculiarities of *ciertamente* and the adverbial phrase *por cierto*, see Estellés (2009).

the framework of Functional Grammar, these adverbs would be considered *proposition satellites* because they take scope over the whole sentence. This category is not very different from Greenbaum's *attitudinal disjuncts*, which is situated in the third layering hosting level, i.e. they are σ_3 satellites. By contrast, Ramat and Ricca (1998, p. 190) consider these types of adverbs as σ_2 satellites because they propose a classification for sentence adverbs alone. These authors consider that σ_2 satellites can be related with "the propositional content of the sentence", so we can see that this term is equivalent to Dik's "proposition satellites".

Notice that the modal meaning of these sentence adverbs is highly subjective (in the sense that it expresses the speaker's involvement in the following statement):

- (6) *Pero, dexando esto, digo que el fundamento, y casi la fuente donde nacen las gracias que hacen reír, consiste en una cierta desproporción o deformidad, si quisieredes así llamalla; porque solamente nos reímos de aquellas cosas que en sí desconviene y parece que están mal, pero realmente no lo están*
 'I affirm that the source of all jokes consists of a certain disproportion or deformity [...]; because we only laugh at those things that seem to be wrong, but which really are not' (Juan Boscán, *Traducción de "El cortesano" de Baltasar de Castiglione, 1534; CORDE*)
- (7) *la quinta es ser señor de buena gente y que lo amen verdaderamente, que si verdaderamente no lo amaren, no puede ser bien servido dellos*
 'The fifth [recommendation] is to be the master of good people and that they truly love him, because if in fact/really they do not love him, he cannot be served well by them'
 (Anónimo, *Libro del cavallero Zifar, 1300–1305; CORDE*)

As sentence adverbs, these forms no longer express a manner of executing the verbal action, but rather the speaker's reinforcement of the sentence he is going to pronounce. The etymological meaning, nonetheless, has not completely disappeared in the examples above. As Hopper and Traugott (2003, p. 97) state, "persistence of old meanings is a common phenomenon", as we can see especially in (7), where *verdaderamente* can still be read as modifying the VP *amaren*, given that the paraphrase 'to love someone in a true way' does not seem so distant.

Finally, we have to point out the relevance of the sentence adverb function in the historical evolution of *verdaderamente*, *realmente* and *ciertamente*. When these adverbs come from functioning as verbal adverbs to widen their structural scope, the following changes take place:

1. A change in their syntactic position: as verbal adverbs, *verdaderamente*, *realmente* and *ciertamente* tend to appear in a postverbal position. On the contrary, as sentence adverbs, they tend to appear in a preverbal position. This happens because they stop modifying the VP and come to occupy a more peripheral position in the statement. This is a key evolution because without

having previously moved to a preverbal position none of these adverbs could have developed a discourse marker function.

2. A change in their structural scope: as sentence adverbs, these terms widen their structural scope in order to modify the whole sentence's proposition content.
3. A semantic change: as sentence adverbs, they no longer express logical notions related to epistemicity, but subjective assessments instead – highly productive semantic content in the field of pragmatics.

5. *Verdaderamente, realmente* and *ciertamente* as discourse markers

When the epistemic adverbs *verdaderamente*, *realmente* and *ciertamente* function as sentence adverbs, it is sometimes difficult to decide whether they take scope over the whole statement and still belong to the propositional level, or if they belong to an extra-propositional level. These ambiguous contexts permit two readings, one interpreting the epistemic adverb as a sentence adverb, and another one interpreting it as a discourse marker, as we can see in example (8):

- (8) *Y ella, con poco entendimiento, dixo que esso mesmo fazia ella a él, mas que no lo osava dezir. e era así, ca ciertamente ella lo amava*
 'And she, with little understanding, said that that was the same thing she was doing to him, but that she did not dare to recognise it, and that it was so, because she truly loved him'

(Anónimo, *Libro del cavallero Zifar*, 1300–1305; CORDE)

If we consider that *ciertamente* in (8) is a sentence adverb we will imply that despite the fact that it expresses speaker's certainty it takes scope over the propositional content of the whole statement *ella lo amava*. If, on the contrary, we read it as a discourse marker, we would have to say that *ciertamente*'s reinforcing meaning is situated at a higher level than the propositional content.

Once the terms analysed have widened their structural scope, they acquire the possibility of appearing in an even more peripheral position as discourse markers. In this function, they exhibit wider structural scope and more autonomous syntactic behaviour, as we can see in the following examples:

- (9) *Oso decirte [...] que si tus pecados fuesen sin número, y ninguno [...] hubiese de quedar sin castigo, como realmente no ha de quedar, [...], digo, [...] que en brevísimo tiempo, y quizá de una hora, [...] excusases penas tan grandes cuanto ninguno puede encarecer ni imaginar*

'I dare to say [...] that if your sins were innumerable, and none [...] was left without punishment, as in fact/really none of them should be, [...] I say [...] that in a very short time period, and maybe in an hour, [...] you would suffer unbelievable pain'

(Fray Juan de los Ángeles, *Diálogos de la conquista del reino de Dios*, 1595; CORDE)

- (10) *e levantar se á este pueblo [...] allí me desamparará [...] e desamparar le é como él a mí, e asconder le é la mi faz [...], e vernán sobr'él todos males e todas pestilencias que seer podieren, de guisa que dizrá él esse día: **verdaderamente** porque non es Dios comigo por ende me vienen todos estos males*

'And this people will rise up [...] and there they will abandon me [...] and I will abandon them like they abandoned me and I will hide my face from them [...] and they see upon them all evils and all pestilences that exists; so that on that day people will say: really/truly, because God is not with me, that is why all these bad things are happening to me'

(Alfonso X, *General Estoria. Primera parte*, ca. 1275; CORDE)

Once discourse markers, these adverbs experience an increase in their syntactic abilities so that they can appear in different positions. We can, nonetheless, define the initial position as the one preferred, but, as we can see in (9), a parenthetical position is also possible. The latter represents the highest level of syntactic autonomy because it is separated by intonation breaks from the rest of the statement.

They also experience a semantic change, as subjective meanings that appear in the sentence adverb function consolidate and shift towards even more pragmatic nuances. As we can see in examples (9) and (10), *realmente* and *verdaderamente* express reinforcement and introduce the speaker's belief about what he is going to say. The relation with etymological meanings is never completely erased, as linguistic change does not necessarily involve loss of original meanings. We can consider the semantic evolution of *verdaderamente*, *realmente* and *ciertamente* as instances of subjectification:

the mechanism whereby meanings come over time to encode or externalize the SP/W's perspectives and attitudes as constrained by the communicative world of speech event, rather than by the so-called 'real-world' characteristics of the event or situation referred to. (Traugott, 2003a, p. 125)

As we have already mentioned, the discourse markers *verdaderamente*, *realmente* and *ciertamente* are more oriented towards the expression of the speaker's beliefs than towards any notion of modality related to the world of facts. We can also see

that the evolution drawn above fits the main features described in Company's definition of the term "subjectification" (2004, p. 41), which include:

1. Weakening and semantic bleaching of the original, etymological referential meaning;
2. Metaphoric-metonymic change of the discourse-pragmatic inferential nature;
3. Increase in scope of the predication;
4. Loss of syntactic properties.

As we have seen, the consolidation of the change whereby adverbs analysed became discourse markers fulfils the first and the third feature. The fourth one refers to the fact that, by virtue of being discourse markers, these adverbs lose some of their syntactic properties (above all, the ability to accept modifiers). The second one refers to the semantic change experienced by *verdaderamente*, *realmente* and *ciertamente*. Company (2004) understands, following the explanation by Traugott and Dasher (2002), that a term becomes more subjective through the conventionalisation of *invited inferences* (see Traugott & Dasher, 2002, p. 35 and ff. for further details). In our case, we saw that reinforcement and emphatic notions are present in some contexts where the adverbs of manner *verdaderamente*, *realmente* and *ciertamente* also surface. With time, these nuances, which could be inferred by the addressee in some contexts, conventionalise and eventually become the new coded meaning of these adverbs as they come to function as discourse markers.

Finally, the linguistic changes described above can be understood as the result of a grammaticalisation process since all the adverbs experience the following micro-changes:

1. They change their grammatical status; this "deategorialisation" can be seen as a change from a relatively major category (adverbs) to that of discourse markers.
2. They increase their structural scope, given that as discourse markers they have a scope over the whole utterance.
3. Their meaning changes from a more specific one, equivalent to 'in a truly, physical, material way', to a more abstract one, namely, the expression of the certainty the speaker has about what is going to say.
4. Therefore, these adverbs undergo a process of subjectification.
5. From a syntactic perspective, their mobility in discourse increases, instead of becoming more rigid.

Even though it is undeniable that pragmaticalisation processes are not prototypical instances of grammaticalisation, some authors have argued that they are

still grammaticalisation processes, albeit characterised by different features (see Wischer, 2000, p. 355 and ff.). In contrast, other authors, for instance, Ajmer (1997), support pragmaticalisation to be a different process from grammaticalisation. Despite the differences between the traditional grammaticalisation definition and pragmaticalisation, in the present paper we agree with the theoretical position adopted by Traugott (1995) and Brinton (2007). The latter states that “pragmaticalization, it seems to me, is just a subspecies of grammaticalization, not a distinct process” (Brinton, 2007, p. 63).¹⁵

6. The development of a new adverbial function: *Verdaderamente, realmente* and *ciertamente* as intensifying adverbs

In sections 3 to 5 we have explained the first grammaticalisation process that *verdaderamente*, *realmente* and *ciertamente* undergo. In this first step, we have seen that these adverbs shift from a VP structural scope to a wider one, and from a more grammatical function to a more pragmatic one.¹⁶ In this section, however, we focus on a second change that some epistemic adverbs experience. In this case, the epistemic adverbs *verdaderamente*, *realmente* and *ciertamente* develop a new adverbial function, but this time as adjective modifiers, in contexts like (11):

- (11) *aquellos omes que tienen vigor de juyzio e entendimiento, e virtud, éstos son verdaderamente libres*
 ‘Those men who have judgment, understanding and virtue, those are truly free’
 (Rodrigo Sánchez de Arévalo, *Vergel de los príncipes*, 1454–1457; CORDE)

We can see in (11) that these adverbs can take scope over the adjective they modify. They also change their semantics: as an adjective modifier, *verdaderamente* shows a mere emphatic meaning. It is still related to the discourse marker meaning, but this has been relatively bleached by the intensifying adverb.

15. Further debate about the definition of grammaticalisation and pragmaticalisation is needed in the area of historical linguistics. However, this issue is beyond the aim of the present paper and should be left aside in order to focus on the specific changes we are describing.

16. We cannot, nonetheless, confirm that adverbs function as purely grammatical items. They are, indeed, more grammatical than other items like verbs or nouns, but not as grammatical as morphemes, for example. Lexical and grammatical poles have not always been clearly defined; hence, we consider them as a gradual scale. Therefore, we can define adverbs as more grammatical than discourse markers, but not as completely grammatical. For further discussion about what “grammatical functions” and “lexical items” mean, see Traugott (2003b, p. 2).

In the following sections we show how this new function emerges in some particular contexts and how, with time, it increases its frequency of use and its syntactic abilities.

6.1 Chronology and distribution

If we look at the first testimonies of epistemic adverbs functioning as adjective modifiers in CORDE, we see that *verdaderamente* was, among the three items that have been analysed here, the first one to show this possibility. In fact, the first examples were found during the second half of the 15th c.

Furthermore, if we observe which lexical categories *verdaderamente* usually modifies throughout history, we can see that, whereas until the 16th c. *verdaderamente* basically modified verbs, from the 17th c. it functions primarily as an adjective modifier. This can be seen in Table 6, where extra-propositional examples have been excluded:¹⁷

Table 6. *Verdaderamente* modifying lexical categories

	13th–14th c.	15th–16th c.	17th–18th c.	19th–20th c.
Adjective	0%	19% (154/826)	53% (177/334)	76% (1122/1485)
Adverb	0%	< 1% (5/826)	< 1% (1/334)	1% (20/1485)
Noun/Noun Phrase	0%	< 1% (4/826)	2% (8/334)	1% (17/1485)
Verb	99% (106/107)	80% (663/826)	44% (148/334)	22% (326/1485)

When considering *verdaderamente* only with verbal structural scope, it seems obvious that this adverb shifted from modifying verbs to adjectives and finally became reanalysed as an adjective modifier with an intensifying meaning.

Using CREA data, we can confirm this last function to be the core meaning of *verdaderamente* also during the 20th and 21st c. Thus, Table 7 confirms that during this last period *verdaderamente* shows verbal structural scope in 76% of the examples:

17. Contexts where *verdaderamente* functions as a discourse marker or as a sentence adverb have thus been included.

Table 7. The scope of *verdaderamente* in the 20th–21st c.

	20th–21st c. (CREA)
Verbal structural scope	76% (704/922)
Non-verbal structural scope	24% (218/922)

Importantly, while analysing CORDE data, we have not been able to find such a clear consolidation of this new function in the case of *ciertamente* and *realmente*. In fact, the first example of *ciertamente* modifying an adjective is found in the 16th c. and, even in the 20th c., this use has not become as frequent as in the case of *verdaderamente*. Nevertheless, we can observe an increase in the use of *ciertamente* as adjective modifier in Table 8:

Table 8. *Ciertamente* modifying lexical categories

	13th–14th c.	15th–16th c.	17th–18th c.	19th–20th c.
Adjective	0%	3% (5/192)	22% (26/120)	31% (137/441)
Adverb	0%	2% (4/192)	5% (6/120)	4% (18/441)
Noun/Noun Phrase	0%	0%	0%	<1% (2/441)
Verb	100% (76/76)	95% (183/192)	73% (88/120)	64% (284/441)

As we can see, the frequency of use of the adjective modifier clearly increases between the 13th and the 20th c. (from 0% to 31%), but *ciertamente* functions as a discourse marker (in 60% of the examples) during the 19th–20th c.

If we look at CREA data, we can see that during the 20th–21st c., *ciertamente* takes non-verbal structural scope (that is to say, either as a sentence adverb or as a discourse marker):

Table 9. The scope of *ciertamente* in the 20th–21st c.

	20th–21st c. (CREA)
Verbal structural scope	35% (286/826)
Non-verbal structural scope	65% (540/826)

Finally, *realmente* shows an evolution similar to *ciertamente*. Although it functions primarily as a verbal adverb, even during the 19th–20th c., its use as an adjective modifier increases its frequency (from 0% to 38% during the 19th and 20th c.) mostly from the 17th–18th c.:

Table 10. *Realmente* modifying lexical categories

	13th–14th c.	15th–16th c.	17th–18th c.	19th–20th c.
Adjective	0%	2% (5/311)	18% (20/111)	38% (312/824)
Adverb	0%	0%	3% (3/111)	1% (9/824)
Noun/Noun Phrase	0%	0%	2% (2/111)	<1% (3/824)
Verb	100% (2/2)	98% (306/311)	77% (86/111)	61% (500/824)

The verbal adverb remains *realmente*'s main function even during the last period. As we can see in Table 11, even when looking at CREA data, we can see that *realmente* shows a preference for functioning with verbal structural scope:

Table 11. The scope of *realmente* in the 20th–21st c.

	20th–21st c. (CREA)
Verbal structural scope	64% (528/826)
Non-verbal structural scope	36% (298/826)

Looking at the functions of *realmente* with verbal structural scope, we can compare the frequency of use as verbal adverb with that of the intensifying adverb.

Table 12. The functions of *realmente* in the 20th–21st c.

	20th–21st c. (CREA)
Intensifying adverb	47% (250/528)
Verbal adverb	53% (278/528)

The former is the most frequent one during the 20th–21st c., although the latter is confirmed as the second most important function (see Table 12).

In conclusion, we have seen that the three adverbs analysed in this paper all confirm the development of a new adverbial function. Among these three adverbs, *verdaderamente* is the first one to show this possibility and the best example of this extension of adverb functionalisation; namely, its reanalysis as an adjective modifier. The best proof of this is that, during the 20th–21st c., its distribution has narrowed and it functions mainly as an intensifying adverb. *Ciertamente* and *realmente*, however, although experiencing a relevant increase in their use as intensifying adverbs, function respectively as a discourse marker or as a verbal adverb.

6.2 Extension of the adjective modifier function

In order to understand how the discourse markers *verdaderamente*, *realmente* and *ciertamente* became intensifying adverbs, we have to bear in mind the following theoretical assumptions:

1. Grammaticalisation is a gradual process: it begins in very specific contexts, semantically similar to the item that undergoes grammaticalisation, and once the new grammatical meaning is consolidated, it spreads to more general contexts.
2. “any category in a language constitutes a *continuum* with prototype effects: some members are focal or typical [...], they are the best instances of the category, they are the prototypes [...]; other members are less typical and yet others are situated on the borderline area of the category, exhibiting grammatical properties of two or more categories” (Company, 2002, p. 201).
3. In every grammaticalisation process, we can isolate critical contexts where two readings are permitted, that is to say, where pragmatic ambiguities emerge.¹⁸ These contexts can be defined as the *locus* of semantic change.

In the case of the epistemic adverbs analysed, we can define contexts in which adverbs appear before a past participle as critical, as in (12):

- (12) *que las cosas verdaderamente amadas cuanto más dellas se alcanza mucho más el desseo y cuidado se aumenta y cresce*
 ‘[...] because the more one gets the truly loved things, the more one desires them’

(Garcí Rodríguez de Montalvo, *Amadís de Gaula*,
 libros I y II, 1482–1492; CORDE)

18. We use the term “critical context” with the same meaning as “bridging contexts” (Heine, 2002) or as “untypical contexts” (Diewald, 2006). Both authors consider this kind of contexts as the essential mechanism by which invited inferences emerge.

That these contexts are critical derives from various facts. Firstly, since past participles are hard to define due to verbal-adjectival category ambiguity, it seems logical to think that this change begins with them. Concretely, they exhibit characteristics from both categories, so it seems plausible to assert that, before becoming adjective modifiers, these epistemic adverbs co-occurred with past participles and, in these contexts, showed a function halfway between verbal adverbs and adjective modifiers. This confirms the hypothesis that categories are not homogeneous and that precisely those elements situated on the borderline between categories are the most prone to experiencing changes.

Secondly, from the semantic perspective, these contexts permit two readings. One can understand that *the really loved things* are the things one loves ‘in a real, physical manner’, or that they are *the very loved things*, in other words, ‘the things that one loves with great intensity’. This is how the intensifying/quantifying notion appears. Once again, this change can be explained through the *invited inferences* model (Traugott & Dasher, 2002). In the first place, the intensifying notion was only a pragmatic nuance the hearer could infer in a specific context. Later on, this inference conventionalises and becomes part of the coded meaning of these epistemic adverbs.

One of the factors that could have motivated this second process of grammaticalisation was the frequent appearance of these adverbs before past participles. In the case of *verdaderamente*, this increases from 7% during the first period corresponding to the 13th and 14th c. to 13% during the 17th–18th c.

We can also verify this same evolution in the case of *realmente*, given the fact that *realmente* increases its frequency of use with past participles from 4% during the 15th–16th c. to 8% during the 19th–20th c. The evolution of *ciertamente* is, nonetheless, more complex. One would expect that the token [adverb + past participle] predates *ciertamente* as an intensifying adverb. However, CORDE data contradict this hypothesis, given the fact that examples where *ciertamente* modifies nouns or adjectives with an intensifying meaning have been documented earlier and with a higher frequency.

We have defined grammaticalisation as a gradual process. In our case, this gradualness is related to the extension of the adjective modifier function. Thus, in the next stage *verdaderamente*, *realmente* and *ciertamente* co-occur with adjectives, given the grammatical similarity between past participles and the adjectival category. However, to start with, they only modify adjectives that express gradual qualities, with a negative or positive polarity, like *poor*, *rich*, *good*, *bad*, *beautiful* or *ugly*, as in the following examples:

- (13) *cada uno quiere y desea, no lo que en sí mismo es verdaderamente bueno, porque tiene el corazón sano, limpio y alumbrado*

‘Each one wants and desires, not what is truly good, because it has a healthy and clean heart’

(Fray Hernando de Talavera, *Católica impugnación del herético libelo maldito y descomulgado*, 1487; CORDE)

The adjectives that we find in examples like (13) denote qualities ranging from highest to lowest. This facilitates their co-occurrence with epistemic adverbs, as they are elements that can be evaluated, and therefore, emphasised.

From these more prototypical contexts, intensifying adverbs reach less prototypical ones, and thus can appear before superlative adjectives.¹⁹ In this case, these epistemic adverbs do not quantify a quality, but rather emphasise one that is positive or negative by definition:

- (14) *por la tarde habían circulado repetidos anuncios de sangrientas escenas, y Madrid estaba realmente aterrado*

‘In the afternoon the press had reported some violent scenes, and Madrid was really terrified’

(José Selgas y Carrasco, *Un rostro y un alma*, 1874; CORDE)

- (15) *Si él puede ir por el atajo y tiene en la mano un medio cierto para librarse de la persecución de los exorcistas que es simular y disimular con ellos, ¿para qué recurre a un medio dudoso y aun ciertamente inútil?*

‘If he [the devil] can escape from the persecution of the exorcists, why should he use such an uncertain and certainly useless mean?’

(Benito Jerónimo Feijoo, *Theatro Crítico Universal*, 1733; CORDE)

Finally, the adverbs analysed co-occur with adjectives, which express a neutral quality, that is to say, a quality that cannot be situated in a positive-negative scale, as in (16):

- (16) *Estando en casa de su padre, se le junto vna gran sierua de Dios matrona verdaderamente biuda, llamauase doña Mayor Gomez de gran espíritu*

‘Staying in his father’s house, a servant of God came to help him, a real widow, whose name was doña Mayor Gomez’

(Fray José Sigüenza, *Segunda parte de la Historia de la Orden de San Jerónimo*, 1600; CORDE)

19. Although *superlativo* in Spanish refers to a very restricted group of adjectives, which share the morpheme *-ísimo*, we have used the term “superlative” to designate a number of adjectives, which, although not showing this specific morphology, semantically express the highest or a very high degree of a quality.

- (17) *el primero de mis amigos, el Licenciado Pedro Pantoja de Ayala, ingenio verdaderamente toledano, [...]*

‘The first of my friends, the graduate Pedro Pantoja de Ayala, a truly Toledan wit [...]

(Tomás Tamayo de Vargas, *Comentarios a Garcilaso*, 1622; CORDE)

As we can see, *biuda* or *toledano* do not show any positive or negative orientation. Neither are concepts that permit quantification, as they designate binary qualities (that is, one cannot be more or less widow). However, from the moment that an epistemic adverb modifies these adjectives, the addressee/reader situates these qualities on a scale and therefore accepts that, in fact, one can be more or less widow or Toledan. It is the presence of the adverb, though, and not the semantics of the adjective, that facilitates this reading.

In conclusion, we can describe the extension of *realmente*, *verdaderamente*, and *ciertamente* as adjective modifiers as following a process that goes from more prototypical to less prototypical contexts:

scalar adjectives > superlative adjectives > neutral adjectives
 [+prototypical contexts] > [+/- prototypical] contexts > [- prototypical] contexts

This is exactly the extension path followed by *verdaderamente*, as we can see in Table 13:

Table 13. Semantics of the adjectives that co-occur with *verdaderamente*

	13th–14th c.	15th–16th c.	17th–18th c.	19th–20th c.
Scalar	0%	53% (79/150)	55% (93/168)	51% (568/1109)
Superlative	0%	29% (44/150)	32% (54/168)	25% (278/1109)
Neutral	0%	18% (27/150)	13% (21/168)	24% (263/1109)

As we can observe, during the 15th and 16th c. the adjective modifier tends to co-occur mainly with scalar adjectives. Even though *verdaderamente* appears with this type of adverb in approximately half of the examples, co-occurrence with superlative adjectives increases between the 17th and the 18th c., and with neutral adjectives it increases between the 19th and the 20th c. (from 18% to 24%).

Adjective modifier *ciertamente* shows a similar evolution, although this adverb was not documented with an intensifying meaning until the 16th c. The co-occurrence with scalar adjectives is the most frequent throughout history:

Table 14. Semantics of the adjectives that co-occur with *ciertamente*

	13th–14th c.	15th–16th c.	17th–18th c.	19th–20th c.
Scalar	0%	60% (3/5)	48% (10/21)	60% (78/132)
Superlative	0%	20% (1/5)	28% (6/21)	26% (35/132)
Neutral	0%	20% (1/5)	24% (5/21)	14% (19/132)

The frequencies of co-occurrence with superlative and neutral adjectives remain steady throughout history (even though neutral adjectives decrease from 20% to 14%).

In the case of *realmente*, given the scarcity of examples that have been encountered, it is difficult to know whether the evolution experienced is different from the others described above or whether this particular behaviour is caused by the lack of examples. In Table 15, we can see that until the last period analysed (19th–20th c.), only 25 examples of the adjective modifier *realmente* were found in CORDE:

Table 15. Semantics of the adjectives that co-occur with *realmente*

	13th–14th c.	15th–16th c.	17th–18th c.	19th–20th c.
Scalar	0%	60% (3/5)	25% (5/20)	34% (106/310)
Superlative	0%	0%	0%	31% (95/310)
Neutral	0%	40% (2/5)	75% (15/20)	35% (109/310)

Twenty-five examples are not enough to establish the patterns of co-occurrence of *realmente*, so we cannot reach definite conclusions for the 13th–18th c. As for the 19th–20th c., we can conclude that *realmente* appears equally frequently with scalar, superlative and neutral adjectives and, therefore, it is very difficult to conclude as to with which of these this last functionalisation began.

Once these adverbs have consolidated an intensifying meaning, they can modify other lexical categories, such as nouns, pronouns or adverbs, i.e. all categories, which are very distant from the prototype (that is to say, scalar adjectives):

- (18) *Es sólido, es conciso, y de gran provecho: y aun aquellas, que solo presidió, y o son realmente suyas, tienen mérito bastante*

‘[Linneo’s work] is solid, brief and useful: and even those works that he only coordinated, as well as those that are really his, show relevant merits’

(Anónimo, *Extractos de las Juntas Generales celebradas por la Real Sociedad Bascongada de los Amigos del País*, 1793; CORDE)

- (19) *He aquí la razón por que se ha escrito con más libertad e independencia en épocas ciertamente mucho más atrasadas que las que nosotros hemos alcanzado*

‘This is the reason why people used to write with more freedom and independence in certainly much more retrograde periods than the one in which we are living’

(Mariano José de Larra, *Representación de “La mojegata” comedia de don Leandro Fernández de Moratín* [Fíguro], 1834; CORDE)

This kind of example is not representative of the adjective modifier function. Nonetheless, it is important to highlight that, eventually, the distribution of intensifying adverbs *verdaderamente*, *realmente* and *ciertamente*, having initially developed this new meaning in connection with past participles and scalar adjectives, has since reached distant contexts, such as the ones formed by nouns, pronouns and adverbs (although not any adverb can co-occur with epistemic adverbs. It must show a quantifying meaning, as in (19) *mucho*. Thus, the gradual process yields from more prototypical, semantically similar contexts to less prototypical, semantically different ones.

7. Conclusions

In sections 3 to 5, we have seen that the verbal adverbs *verdaderamente*, *realmente* and *ciertamente* grammaticalised as discourse markers in a grammaticalisation process that goes from grammar to discourse. This process can be understood as a pragmatization process, as it leads epistemic adverbs from the lexicon to the discourse level. As we said in Section 5, many authors have defined these evolutions as a special subtype within the grammaticalisation processes.

In Section 6, we have shown that, later on, these discourse markers experience a new functionalisation as adjective modifiers. We can define this evolution as a “double movement of meaning extension” (Willems & Demol, 2006, p. 220),²⁰ given that the

20. See Willems & Demol (2006) for a detailed description of the historical evolution of French epistemic adverb *vraiment* and of the English adverb *really*. These adverbs show similar

core, original meaning of these epistemic adverbs firstly shifts into a more pragmatic, intersubjective one and then becomes a discourse marker, and secondly develops a new intensifying meaning, which is consolidated by the adjective modifier function.

Finally, we have to remark that all the functions described in this paper co-exist and remain in Spanish today. With time, the manner adverb function has decreased its use, but it still exists in the language. The development of different meanings has lead epistemic adverbs to show a high degree of poly-functionality and of layering (Hopper, 1991), a common phenomenon in grammaticalisation processes.

References

- Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
- Blakemore, D. (1992). *Understanding utterances*. Oxford: Blackwell.
- Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486456
- Brinton, L. (2007). The development of *I mean*: Implications for historical pragmatics. In S. M. Fitzmaurice & I. Taavitsainen (Eds.), *Methods in historical pragmatics* (pp. 37–79). Berlin: Mouton de Gruyter. doi:10.1515/9783110197822.37
- Company, C. (2002). Grammaticalization and category weakness. In I. Wischer & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 201–123). Amsterdam: John Benjamins. doi:10.1075/tsl.49.14com
- Company, C. (2004). ¿Gramaticalización o desgramaticalización? Reanálisis y subjetivización de verbos de movimiento como marcadores discursivos en la historia del español. *Revista de Filología Española*, 54, 29–66. doi:10.3989/rfe.2004.v84.i1.97
- Corominas, J., & Pascual, J. A. (1991). *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos.
- Davies, M. (2008). Creating useful historical corpora: A comparison of *CORDE*, the *Corpus del español* and the *Corpus do português*. In A. Enrique-Arias (Ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus* (pp. 137–167). Madrid & Frankfurt: Iberoamericana & Vervuert.
- Diewald, G. (2006). Context types in grammaticalization as constructions. In D. Schöenfeld (Ed.), *Constructions all over: Case Studies and theoretical implications, Constructions SVI*, 9, 1–29.
- Dik, S. C., Hengeveld, K., Vester, E., & Vet, C. (1990). The hierarchical structure of the clause and the typology of adverbial satellites. In J. Nuyts, A. M. Bolkestein & C. Vet (Eds.), *Layers and levels of representation in language view. A functional view* (pp. 25–70). Amsterdam: John Benjamins. doi:10.1075/pbns.13.04dik
- Dik, S. (1997). *The theory of Functional Grammar. Part 1: The structure of the clause*. Berlin: Mouton de Gruyter.
- Egea, E. R. (1979). *Los adverbios en -mente en el español contemporáneo*. Bogotá: Instituto Caro y Cuervo.

semantic changes to the ones described in the present paper for Spanish adverbs *verdaderamente*, *ciertamente* and *realmente*.

- Estellés, M. (2009). *Gramaticalización y gramaticalizaciones. El caso de los marcadores del discurso de digresión en español*. (Unpublished doctoral dissertation). Universitat de València, Valencia, Spain.
- Greenbaum, S. (1969). *Studies in English adverbials usage*. London: Longman.
- Heine, B. (2002). On the role of context in grammaticalization. In I. Wischer, & G. Diewald (Eds.), *New reflections on grammaticalization* (pp. 83–97). Amsterdam: John Benjamins. doi:10.1075/tsl.49.08hei
- Hengeveld, K. (1989). Layers and operators in Functional Grammar. *Journal of Linguistics*, 25(1), 127–157. doi:10.1017/S0022226700012123
- Hopper, P. J. (1991). On some principles of grammaticization. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization* (pp. 17–36). Amsterdam: John Benjamins. doi:10.1075/tsl.19.1.04hop
- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139165525
- Jackendoff, R. S. (1972). *Semantic interpretations in generative grammar*. Cambridge, MA: MIT Press.
- Karlsson, K. E. (1981). *Syntax and affixation. The evolution of MENTE in Latin and Romance*. Tübingen: Max Niemeyer. doi:10.1515/9783111329017
- Kovacci, O. (1999). El adverbio. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 1; pp. 705–786). Madrid: Espasa.
- Lyons, J. (1977). *Semantics*. London: Cambridge University Press.
- Lleal, C. (2010). El DICCA-XV y el estudio de los procesos evolutivos. *Revista de Historia de la Lengua Española*, 5, 91–103.
- Martín Zorraquino, M. A. (2010). Los marcadores del discurso y su morfología. In O. Loureda Lamas & E. Acín Villa (Eds.), *Los estudios sobre marcadores del discurso en español, hoy* (pp. 93–183). Madrid: Arco/Libros.
- Palmer, F. R. (1986). *Mood and modality*. Cambridge: Cambridge University Press.
- Pountain, Ch. P. (2018). Variation and the use of discourse markers in 16th-century Spanish. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 303–324). Amsterdam: John Benjamins. (in this volume) doi:10.1075/ihll.16.13pou
- Ramat, P., & Ricca, D. (1998). Sentence adverbs in the languages of Europe. In J. van der Auwera (Ed.), *Adverbial constructions in the languages of Europe* (pp. 187–275). Berlin: Mouton de Gruyter. doi:10.1515/9783110802610.187
- Rodríguez Ramalle, T. M. (2003). *La gramática de los adverbios en -mente o cómo expresar maneras, opiniones y actividades a través de la lengua*. Madrid: Ediciones de la Universidad Autónoma de Madrid.
- Traugott, E. C. (1989). On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language*, 57, 33–65.
- Traugott, E. C. (1995). The role of the development of discourse markers in a theory of grammaticalization. Paper presented at the *12th International Conference on Historical Linguistics*, University of Manchester, August 1995. Retrieved from <http://www.stanford.edu/~traugott/papers/discourse.pdf>
- Traugott, E. C. (2003a). From subjectification to intersubjectification. In R. Hickey (Ed.), *Motives for language change* (pp. 124–139). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486937.009

- Traugott, E. C. (2003b). Constructions in grammaticalization. In B. Joseph & R. Janda (Eds.), *The handbook of historical linguistics* (pp. 624–648). Malden, MA: Blackwell.
doi:10.1002/9780470756393.ch20
- Traugott, E. C., & Dasher, R. B. (2002). *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Varela, S. (1990). *Fundamentos de morfología*. Madrid: Síntesis.
- Wilson, D., & Sperber, D. (1993). Linguistic form and relevance. *Lingua*, 90, 1–25.
doi:10.1016/0024-3841(93)90058-5
- Willems, D., & Demol, A. (2006). *Vraiment and really* in contrast: When truth and reality meet. In K. Aijmer & A. -M. Simon-Vandenberg (Eds.), *Pragmatic markers in contrast* (pp. 215–237). Amsterdam: Elsevier.
- Wischer, I. (2000). Grammaticalization versus lexicalization. ‘Methinks’ there is some confusion. In O. Fischer, A. Rosenbach & D. Stein (Eds.), *Pathways of change. Grammaticalization in English* (pp. 355–370). Amsterdam: John Benjamins. doi:10.1075/slcs.53.17wis

Corpus

- Real Academia Española. *Corpus diacrónico del español (CORDE)*. Madrid: RAE. <<http://www.rae.es>>
- Real Academia Española. *Corpus del español actual (CREA)*. Madrid: RAE. <<http://www.rae.es>>

Variation and the use of discourse markers in 16th-century Spanish

Christopher J. Pountain
Queen Mary, University of London

This study aims to account for the differing uses and frequency profiles of a variety of discourse markers (DMs) in four roughly contemporaneous 16th-century Spanish texts which are contrasting in nature.¹ Such a methodology complements previous studies which have used large heterogeneous corpora, in that it controls the variables of authorship and text-type as well as allowing for a closer philological appraisal of the data. It is able to show not only that the presence, absence or frequency of a particular DM is determined by the tenor and field of the text, as well as by its characteristic discourse type, but also that when DMs are variant exponents of the same function, they may also act as social indicators.

Keywords: discourse markers; 16th century; Spanish; Lope de Rueda; variation; social indicator

1. Introduction

Surprisingly, given its wealth of text types and its chronological position in what is often perceived as a period of rapid change in the history of Spanish, the 16th century has been relatively little studied in relation to discourse markers (DMs), as Pons Rodríguez (2010, p. 482) has pointed out (see also González Manzano, in this volume, for the historical development of three Spanish epistemic adverbs). Exception must be made of Cano Aguilar's (2007) survey, which focusses on processes of grammaticalisation and degrammaticalisation, and on the arguably

1. I gratefully acknowledge the comments of participants in the Second Cambridge Colloquium on the History of the Ibero-Romance Languages – Norman MacColl Symposium 2012, at which a version of this paper was presented, and of two anonymous referees.

pivotal chronological position of the 16th century between the ‘medieval’ and ‘modern’ language, of Fernández Alcaide (2009), who has paid particular attention to the identification and classification of DMs in her analysis of the language of a corpus of private correspondence from the 16th century, and of Sánchez Jiménez’s (2008) study of DMs in Lope de Rueda. The starting point of the research I present here, however, was in origin not an investigation of DMs *per se*, but an excursus I made from a project which was designed to examine sociolinguistic variation in the prose plays of Lope de Rueda (?–1565), written in the 1540s and 1550s, and published in 1567 (for some other preliminary results of this project, see Pountain, 2009, 2012). As sociolinguistic markers and indicators, DMs may be significant variables in discriminating the speech of characters of different social levels, and it is reasonable to suppose that Rueda might have made stereotypical use of them in something of the same way that he stereotypically represents phonetic features of the speech of black people and gypsies, or the Latinate syntactic features (*retórica*) of the upper-class characters. Unfortunately, however, it turned out that, except in one or two instances, some of which I will refer to here, DMs did not supply a sufficient number of tokens in order for significant conclusions to be drawn simply on the basis of the Rueda plays, and in an attempt to place the Rueda data in a wider context I extended the investigation to three other roughly contemporaneous mid-16th-century texts of different genres: Tomás de Mercado’s (?–1575) *Summa de tratos* (1571), a treatise on business ethics; Santa Teresa’s (1515–82) spontaneously written spiritual journal *Libro de la vida* (1562), and Fray Luis de Granada’s (1504–1588) *Guía de pecadores* (originally written 1556, though the 1567 edition has been used), a series of sermons, all of which were conveniently available in electronic format, which facilitated quantitative analysis.

2. Methodology

A number of recent valuable diachronic studies on individual DMs have been undertaken using the large corpora with which Spanish is well provided (see the extensive bibliography in Pons Rodríguez, 2010). The advantage with this method is that it returns a significant (though even so in some instances not large) number of tokens. However, it is difficult to control the variables of time, place and style: this can only be realistically achieved by observing usage in just one text, or at best in one text type or one author, on the basis of which we might expect to give a more rigorous account of variation according to individual preferences and tendencies, and this is the method I have adopted here.

The size and dates of my author-based corpus were, as I have explained, determined in the first instance by those of the Rueda texts, which consist of the

dialogue from the *Comedias*, the *Pasos* and the pastoral dramas, the size of which is 63,772 words and of course not extendable. Actually, the size of the other three texts examined was slightly larger though kept to below 100,000 words each (*Vida*, chapters 1–30, 76,803 words; *Guía*, 95,095 words; *Suma*, Books I and II, 75,017 words – these figures appear in Table 1). Even with such limitation, each text offers its own philological difficulties and issues concerning consistency of register: for example, the Rueda corpus as a whole has been taken to represent dialogue, despite the many different styles of speech encountered, and *Guía* includes a good deal of directly translated Biblical material. On the whole, however, I propose that we can consider that Rueda gives some insight into the nature of the direct speech of his time; *Vida* is an interesting, possibly unique, example of spontaneous written language from a first-person viewpoint;² *Guía* is what we might regard as planned discourse directed to a large and varied audience, while *Suma* is the most technical and formal register of the four, and was probably never intended as anything but a written text. The texts were initially examined for 101 of the Modern Spanish DMs listed in Martín Zorraquino and Portolés (1999), of which 34 were attested, and to which 16 others were added as a result of identification in the texts themselves.

Whether this is a sufficient target size for research on DMs must, I think, remain an open question. While I am not aware, on the basis of a number of spot checks, of DMs not attested in these samples which are attested in larger samples of the individual authors or even in their complete oeuvre, it is striking that there are many expected DMs and even whole categories of DMs which are not present. For example, the modern particularising operator *por ejemplo*, which is common in Modern Spanish, does not occur as a DM at all in the four texts: in fact, the phrase only occurs once in the fuller construction *pónese por ejemplo*:³

- (1) *Pónese por ejemplo señaladamente la materia de los pleitos* ‘Take as a clear example [...]’ (Guía, I, xviii)

However, *por ejemplo* is encountered as a DM in the 16th century (though not before): the Corpus del Español records 1,556 instances of which 54 appear to function as DMs in just 11 texts; of these, Luis de Molina’s (1535–1600) *Tratado*

2. For a critical assessment of linguistic evaluations of Santa Teresa’s style, see Mancho Duque (2008).

3. A particular presentational problem in giving an account of DMs is that of providing sufficiently contextualised examples, and in what follows I have provided an English translation of just the crucial material in each example, indicating both the Spanish DM and, in an abbreviated gloss, what I take to be its English equivalent in bold; material in square brackets is given for purposes of clarification.

sobre los préstamos y la usura (1568), which is a text apparently similar in genre and of a similar date to *Suma*, accounts for 31.

- (2) *Así, por ejemplo, si el prestamista no quisiera conceder el beneficio del préstamo si ella por su parte no consintiera en darle gusto con esa torpeza*
 ‘So, for instance, [...]’

(Luis de Molina, *Tratado sobre los préstamos y la usura*)

This immediately suggests that the use of a DM may be a matter of individual preference, which in this case is not shared by the four authors I have examined. The particularising function is expressed in *Suma* by 17 instances of the Latinate *verbigracia*, a DM which is totally absent from the other three texts in my study, where particularisation is still achieved by full clauses (*poniendo/pongamos por ejemplo, ejemplo (desto) es*) rather than by a parenthetical DM element.

It is not my intention in this article to provide a comprehensive study of the DMs identified in the four texts; I limit my attention to those categories which have proved most suggestive with regard to variation: commentators (Section 3.1), ordering markers (Section 3.2), and consecutive connectors (Section 3.3).

3. Analysis

In giving the results of the investigation, I will also outline what I see as some of the practical problems in the identification of DMs and their values. One particular difficulty is that although individual lexical items can be easily retrieved automatically from a digitised corpus (for this purpose I used the concordancing program from WordSmith Tools 5), each example must be checked manually for function, since many words which act as DMs have a range of other functions, and, indeed, varying values as DMs. This necessarily raises the well-known general theoretical problem (see, for example, Fraser, 1999) of what should be included in the category of DM, since this notion cannot be purely lexically defined. I have therefore adopted as a broad limiting definition, following Martín Zorraquino and Portolés (1999, p. 4057) and Garcés Gómez (2008, p. 204), that of an invariable element which expresses a relation of coherence between sentences in discourse and has as its scope a whole proposition. I thus exclude such phenomena as adverbials which modify individual constituents of clauses, and anaphoric elements which effectively embed a preceding element as a constituent of a clause; I have also excluded subordinating conjunctions which mark the function of an embedded clause. Thus many DMs are characterised by their parenthetical nature, and are very often (though not always) sentence-initial.

3.1 Commentator ‘pues’

I begin with a typically problematic case. *Pues* (Table 1) is the most frequent of the DMs examined, and the boundaries between its different values are not always easy to establish: Martínez García (1990, p. 608) notes that many contexts in which *pues* appears are ‘semánticamente ambiguos’.⁴

Table 1. Occurrences of ‘commentary’ *pues*

	Rueda (63,772 w.)		Vida (76,803 w.)		Guía (95,095 w.)		Suma (75,017 w.)	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
<i>pues</i> (all occurrences)	371	58.18	195	25.39	361	37.96	126	16.80
<i>pues</i> ('commentary')	265	41.55	56	7.29	28	2.94	1	0.13

I have distinguished three broad categories of usage:

(a) Use as a causal conjunction (Eng. ‘since’), as in example (3), which continues in Modern Spanish, where it introduces an explicative causal clause and cannot be sentence-initial (*NGLE*, § 46.3j–k); in this function it is regarded as a coordinating conjunction (Bosque, 1989, pp. 195–196). In the 16th century, however, it also functioned as a subordinating conjunction, as in (4), from *Vida*, in which the subordinate clause precedes the main clause. Both these usages are excluded from consideration as a DM.

- (3) *Por otra parte, este justo precio no es el que la ropa vale de suyo y es menester buscar otro nuevo, y, con razón, ninguno de los que la ropa tiene vendida de otra manera es el que se ha de mirar en la almoneda, pues este modo de vender es distinto de los demás.*

‘[...] none of [the prices] at which goods have sold in a different way should be considered at an auction, **because** this manner of selling is different from the others’
(*Suma*, II, xii)

4. In Tables 1–11, the first figure given is the total number of occurrences, while the second is the normalised frequency per 10,000 (ptt) words expressed to two decimal places: this measure of frequency has been chosen to avoid the very small numbers which a percentage would in some cases have yielded (0.13 ptt = 0.0013%).

- (4) *Tengo por una de las grandes mercedes que me ha hecho el Señor este ánimo que me dio contra los demonios. Porque andar un alma acobardada y temerosa de nada sino de ofender a Dios, es grandísimo inconveniente. Pues tenemos Rey todopoderoso y tan gran Señor que todo lo puede y a todos sujeta, no hay qué temer, andando – como he dicho – en verdad delante de Su Majestad y con limpia conciencia.*
 ‘[...] Since we have an all-powerful King [...] there is no reason to fear [...]’
 (Vida, 26.1)

- (b) Use in a ‘commentating’ function, introducing a sentence which is an elaboration or explanation of the preceding discourse (this is the ‘*pues reactivo*’ of Cano Aguilar (2005, p. 137) and the ‘*pues fático*’ of Martínez García (1990, p. 600). This is the most difficult value of *pues* to evaluate and to distinguish from (c), as a comparison of examples (5)–(10) and (35)–(41) will show.
- (c) The sentence-initial use of *pues* as a consecutive connector, to be discussed in Section 3.3.1 below.

Two examples of the commentary function (b) are given in (5) and (6). In (5) Ginesa is acceding to Socrato’s request, indicating that she accepts the reasons he gives for it, and in (6) she asks for further information about what has happened to Pablos, for an explanation of how he came to fall, to which he has alluded rather casually.

- (5) SOCRATO. *No quiero por agora, sino por vida vuestra, ama, que os entréis allá dentro, y le aconsejéis de vuestra parte lo que mejor os pareciere, pues veis que le cumple, y aderéçame essa casa, que yo quiero ir a verme con esos señores.*
 GINESA. *Pues yo me entro, señor.*
 ‘[...] Well [in that case], I’ll go in’ (Lope de Rueda, *Camila*)
- (6) PABLOS. [...] *que creo que se me ha mudado el tono de la voz, como la color de los vestidos con la caída que di.*
 GINESA. *Pues ¿cómo caístes, o quién os hizo caer?*
 ‘[...] So [then], how did you fall [...]?’ (Lope de Rueda, *Camila*)

In neither of these cases is *pues* exactly commutable with the consecutive *de modo que* ‘so, therefore’; it rather has the value of *en este caso* ‘in that case; that being the case’, and indicates that the speaker is bearing the preceding discourse in mind. The pragmatic link with both the causal and consecutive functions of *pues* is clear: the speaker is making the comment on the basis of the instruction just received or the information just given. In this commentary function, *pues* is typically sentence-initial and in dialogue signals the beginning of a new turn which elaborates on

the previous one. For this reason, no doubt, Rueda has a significantly elevated number of occurrences of this kind of usage of *pues*, and we can conclude that, in 16th-century Spanish, as in Modern Spanish, this usage is most characteristic of the spoken language: see also Sánchez Jiménez (2008, p. 2175). Conversely, in *Suma*, the text that is perhaps furthest from spoken register, I have identified only one example of commentary *pues* (7):

- (7) [...] *y si alguno dijere que en cinco ni seis años conoce las cosas de ella, engaña-se. Pues en verdad que era uno de los príncipes de mejor ingenio y entendimiento que ha habido en Europa.*
 '[...] For indeed he was one of the most intelligent and wisest princes [...]'
 (*Suma*, II, xviii)

The correlation between the different values of *pues* and register in Modern Spanish is well-known, and it is legitimate to conclude that the clear distinction between Rueda and the other texts I have examined in this respect confirms a similar basis for the distinction in the mid-16th century (see also Iglesias Recuero, 2000, p. 211, who highlights the problem of 'reconstructing' the orality of medieval Castilian).

Vida and *Guía* each display characteristic functions for what can still broadly be termed the commentary *pues*. In *Vida*, it is used at the beginning of a section or chapter in a way that superficially appears to be à propos of nothing, but which in fact serves as a general indication that the ensuing material is indeed coherent with what has gone before. For example, Chapter 7 of *Vida* begins:

- (8) *Pues así comencé, de pasatiempo en pasatiempo, de vanidad en vanidad, de ocasión en ocasión, a meterme tanto en muy grandes ocasiones y andar tan estragada mi alma en muchas vanidades, que ya yo tenía vergüenza de en tan particular amistad como es tratar de oración tornarme a llegar a Dios.*
 'And [so] in this way I began [...]'
 (*Vida*, 7.1)

In *Guía*, the commentary *pues*, again not representing a full logical consequence, is typically used to introduce a parenthetical comment (9) or a continuing comment (10):

- (9) *Lea quien pudiere los Opúsculos de san Buenaventura, que fue un doctor tan señalado en letras, en devoción, en religión, en prudencia de gobernar – pues a los trece años de su profesión fue general de su orden, y después obispo y cardenal –, y ahí verá cuántas maneras de potajes hace este santo de la vida y pasión de Cristo [...]*
 '[...] indeed, thirteen years after taking his vows he was the General of his order [...]'
 (*Guía*, *De doce singulares provechos...*, I)

- (10) [...] *el teólogo escolástico no se contenta con el lugar de en medio, sino pone su silla sobre todos; y a ninguno le faltan razones, y grandes razones, para creer que su ciencia es la mejor y más necesaria. Pues esto que se halla en las ciencias tan descubiertamente, se halla en las virtudes, aunque más disimuladamente [...]*
 '[...] For [indeed] what is found so openly in the sciences is also found in virtues, though more covertly' (Guía, I, xvi)

3.2 Ordering markers

3.2.1 Opening and continuity markers

DMs of order are preponderant in *Suma* and *Guía*. Indeed, those involving specific numerals (*(lo) primero*, *primeramente* and *lo segundo*) are barely attested in *Rueda* and not present at all in *Vida*:

Table 2. Opening and continuity markers

	Rueda		Vida		Guía		Suma	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
<i>lo primero</i>	0	0.00	0	0.00	0	0.00	30	4.00
<i>primero</i>	0	0.00	0	0.00	5	0.53	1	0.13
<i>primeramente</i>	0	0.00	0	0.00	10	1.05	3	0.40
<i>lo uno</i>	1	0.16	0	0.00	0	0.00	16	2.13
<i>lo segundo</i>	0	0.00	0	0.00	3	0.32	28	3.73
<i>lo tercero</i>	0	0.00	0	0.00	6	0.63	10	1.33

Lo primero and *lo segundo* are also used anaphorically in *Suma*, as in (11); these uses have not been counted as DMs.

- (11) [...] *los preceptos del derecho son ser el hombre en sí justo y a nadie injusto. Para lo primero sirven la prudencia, templanza y fortaleza; para lo segundo, la justicia con sus virtudes anexas y consiguientes, de que ahora no es tiempo de tratar.*
 '[...] The former [man's being just in himself] is achieved by prudence, temperance and fortitude; the latter [man's not being unjust to others] by justice, with its attendant virtues [...]' (*Suma*, I, ii)

There is some evidence of essentially lexical variation according to author: in *Guía*, *primeramente* is exclusively preferred (12), while *Suma* opts predominantly for *lo primero* (13), although *lo uno* (14) and *primeramente* are also found in this text (the DM use of *lo uno* is totally absent from *Guía*).

- (12) *Y por servicios necesarios entendemos primeramente la guarda de los mandamientos de Dios, sin la cual no hay salud. Lo segundo, la guarda de los mandamientos de aquellos que están en su lugar [...]*
 '[...] we should understand, **first of all**, keeping God's commandments [...]. **Secondly**, keeping the commandments of those who act on His behalf'
 (Guía, I, xii)
- (13) *Lo primero, es este negocio de estancos tan odioso, que uno que haya en un pueblo le parece a la gente está cautiva; mas, viendo que el provecho es para su república, llévanlo con mejor ánimo. Lo segundo, siendo oficiales públicos, tratan los negocios y exacciones con más blandura y humanidad.*
 'First, this trade in monopolies is so hateful [...]. **Secondly**, since they are public officials [...]'
 (Suma, II, vi)
- (14) *Y, hablando generalmente, muy mejor juzga la ley que el hombre: lo uno, porque la ordenaron muchos sabios, los cuales entienden mejor los negocios que uno; lo segundo, no les movía pasión de amor ni de interés [...]*
 '[...] the law judges much better than man: **first**, because many wise men established it [...]; **secondly**, [because] they were not motivated by passion or self-interest'
 (Suma, I, iii)

Although apparently not immediately significant, the single use of *lo uno* as a DM in Rueda is also worthy of comment. It comes from the speech of Ysacaro, a shepherd in the pastoral drama *Tymbria*. This class of characters typically exhibit upper-class features, as is evident in Ysacaro's turn, which I quote in full (15):

- (15) YSACARO. *Mira, mira, Troyco, si tal imaginas o piensas, sábetete que en la mitad de la cuenta vives engañado, por esso haz de tu voluntad propria, lo que de fuerça serás constreñido de hazer, lo cual cumpliendo, lo uno como dizen seremos de aquí adelante fieles amigos, y lo otro escusarás que tú a mí, o yo a ti nos busquemos con acechanças lo último de la vida.*
 '[...] when this is done, **first**, we will be, as they say, friends from now on and, **secondly**, we will avoid seeking one another's death by stealth'
 (Lope de Rueda, *Tymbria*)

In this example we see clear markers of upper-class rhetorical style: the relative with *cual*, the absolute construction *lo cual cumpliendo*, the preposing of the adjective (*fieles amigos*) and the euphemistic *lo último de la vida* (for *muerte*). *Lo uno* is contraposed to *lo otro*, enumerating the bases of Ysacaro and Troyco's future relationship. Its slightly unexpected use in this text, therefore, is likely to be a feature of this style.

Several continuity markers appear to peak in usage between the 15th and 17th centuries and then decline (Table 3); these are consequently not mentioned in Martín Zorraquino and Portolés (1999). *Ítem*, which indicates another item in a list with equal status to those preceding (16), was probably always restricted

to formal language (Eberenz, 1994, p. 9; Cano Aguilar, 2007, p. 12; Fernández Alcaide, 2009, pp. 143–144), and appears only in *Guía* and *Suma*. *Demás de esto*, which presents an additional item in a list (17) (Cano Aguilar, 2007, p. 8) is similarly limited to these two texts.

- (16) *Ítem mercar cien potros para hacer caballos y, hechos, venderlos en una feria, trato es de escuderos.*
 ‘Next, to buy a hundred colts in order to rear horses [...]’ (*Suma*, II, ii)
- (17) *Y demás de esto, todas las Escrituras claman que Dios enseña a los humildes y que es maestro de los pequeñuelos, y que a ellos comunica sus secretos.*
 ‘And besides this, all the Scriptures claim that God teaches the humble [...]’ (*Guía*, I, x)

Table 3. More continuity markers in *Suma* and *Guía*

	Rueda		Vida		Guía		Suma	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
<i>ítem</i>	0	0.00	0	0.00	11	1.16	24	3.20
<i>demás de esto</i>	0	0.00	0	0.00	1	0.11	17	2.27

3.2.2 ‘Mayormente’ and ‘especialmente’

Mention may also be made here of what may be regarded as the focusing markers *mayormente* (18) and its near-synonym *especialmente* (19), which, like the continuity markers *ítem* and *demás de esto*, appear only in *Guía* and *Suma*.

- (18) *Demás de esto, llevándole más del justo valor porque ha de ganar en ello, siendo la ganancia incierta, ¿cómo llevas luego parte no habiendo el otro nada? Mayormente, no habiendo tú corrido riesgo ninguno, negocio que, dado fueras compañero, era ilícito, cuanto más siendo vendedor.*
 ‘Besides this, [...] **Most importantly**, because you have not run any risk at all, [it is] a business that, even if you were a partner, would be illicit, so all the more so if you are the seller’ (*Suma*, II, xiv)
- (19) *Y especialmente os amonesto que, después de las escrituras santas, os ejercitéis en el estudio de las Colaciones de los Santos Padres [...]*
 ‘**Particularly** I admonish you [...] to exercise yourselves [...]’ (*Guía*, *Tratado enviado al P. Fray Jerónimo de Ferrara*)

While *especialmente* is also found in *Rueda*, it does not seem to be used as a DM in this text (for both *mayormente* and *especialmente* I have distinguished the DM

usage from adverbial usages which have as their scope a single constituent or embedded clause). The total occurrences of these two items in the texts are given in brackets in Table 4.

Table 4. *Mayormente* and *especialmente*

	Rueda		Vida		Guía		Suma	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
<i>mayormente</i> (total)	0	0.00	0	0.00	1 (15)	0.11 (1.58)	2 (19)	0.27 (2.53)
<i>especialmente</i> (total)	0 (15)	2.35	0	0.00	4 (24)	0.42 (2.52)	2 (28)	0.27 (3.73)

We may note that *especialmente* subsequently increased its frequency in Modern Spanish (apparently at the expense of *mayormente*: Figure 1 gives the comparative raw occurrences attested in the *Corpus del español*).

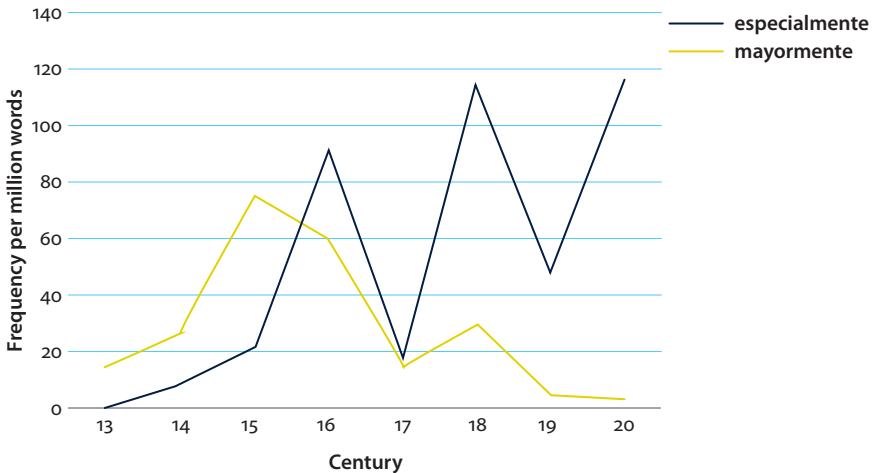


Figure 1. Frequency of *mayormente* and *especialmente* (total occurrences) in the *Corpus del español*

3.2.3 Spatial order markers

Spatial order markers have significant attestation in *Suma*, *Guía* and *Vida* (some examples are given in (20–22)), but are absent from *Rueda*:

Table 5. Spatial order markers

	Rueda		Vida		Guía		Suma	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
<i>por una parte</i>	0	0.00	5	0.95	9	0.95	3	0.40
<i>en parte</i>	0	0.00	3	0.39	0	0.00	2	0.27
<i>por otra parte</i>	0	0.00	7	0.91	1	0.11	6	0.80
<i>por otra [parte]</i>	0	0.00	2	0.26	9	0.95	1	0.13

- (20) *Por una parte me llamaba Dios; por otra, yo seguía al mundo.*
 ‘On the one hand [...]; on the other [...]’ (Vida, 7.17)
- (21) *Es como si estuviésemos hablando con uno y por otra parte nos hablase otra persona [...]*
 ‘It is as if we were speaking to someone and [yet] on the other hand another person were speaking to us’ (Vida, 17.4)
- (22) *Esto, aunque parece todo uno, es diferente de la oración de quietud que dije, en parte, porque allí está el alma que no se quería bullir ni menear [...]*
 ‘Although this appears the same, it is different from the quiet prayer I described; in part, because in that the soul would not try to bubble or shake’ (Vida, 17.4)

Yet even here there are differences of detail. *Guía* makes significant use of contrastive *por una parte... por otra...*; and this function with clauses (23) is essentially no different from that with individual nouns (24):

- (23) [...] *pues nos consta que estos merecimientos, por una parte son tan grandes que no pueden ser mayores, y por otra son tan nuestros [...]* como si nosotros mismos los hubiéramos ganado por nuestra lanza.
 ‘for we find that these merits are on the one hand so great that they cannot be any greater, and on the other belong to us just as is we ourselves had won them with our lance’ (Guía, I, xii)
- (24) [...] *pesando en una balanza, por una parte todas estas pérdidas, y por otra el interés y golosina del pecado [...]*
 ‘[...] weighing in a balance, on the one hand all these losses, and on the other the self-interest and attractiveness of sin’ (Guía, I, iii)

By contrast, there is just one example of *por otra parte* in *Guía*, where it is used as an independent DM with no corresponding contrastive *por una parte* (25); in *Vida*

this latter usage is more frequent (26), to the extent that there must be some doubt as to whether it is really being used as an order marker at all.⁵

- (25) *Si por otra parte dices que es grande la misericordia de Dios [...]*
 'If, **on the other hand**, you say [...]' (Guía, I, viii)
- (26) *Pena que, por otra parte conforta de ver cuán gran merced la hace Dios [...]*
 '[A source of] distress which, **on the other hand**, comforts [us] when we see what great favour God does' (i.e. 'on the one hand, it is a source of distress, but, on the other, a source of comfort') (Vida, 30.9)

In general, then, all the foregoing DMs of order appear to be largely characteristic of dialectic or didactic prose, or are associated with cultured speakers.

3.2.4 Other continuity markers

The most general DM of continuity,⁶ *asimismo* 'in the same way' (Cano Aguilar, 2007, pp. 10–11), is significantly attested in Rueda, but is limited to higher class speakers (the shepherds Asobrio, Troyco and the harpy Mesiflua (27) in *Tymbria*, and Armelina in *Armelina*); it is also present in *Guía* (Table 6). However, it is not attested at all in the sections of *Suma* and *Vida* I examined (though with regard to *Suma*, this may be fortuitous, since there are instances in the later parts of this work which were not included in the selected sample). Overall, therefore, despite the raw impression of association with dialogue, *asimismo* was probably more typical of written registers.

Table 6. Continuity marker *asimismo*

	Rueda		Vida		Guía		Suma	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
<i>asimismo/así mismo</i>	7	1.10	0	0.00	3	0.32	0	0.00

5. See Cano Aguilar (2007, p. 20), who observes that *por otra parte* on its own is 'un conector más de adición de contenidos', and Fernández Alcaide (2009, p. 175), who finds that *por otra parte* is sometimes used simply to indicate a change of topic.

6. Cano Aguilar (2007, pp. 9–10) finds *otosí* to have substantial attestation as a continuity marker, but the texts I examined yielded only three examples of this DM (2 in *Suma* and 1 in *Guía*).

- (27) MESIFLUA. *Y que así mismo el señor Asobrio recompense a Urbana mi sobrina el amistad que sin conocerse se han tenido [...]*
 'And similarly that Señor Asobrio should repay my niece Urbana the friendship which they had without knowing it' (Rueda, *Tymbria*)

Luego, on the other hand, displays a broadly similar frequency in all four texts. The DM function of logical consequence and the purely sequential temporal adverbial function of this element are difficult to distinguish (28):⁷ the statistics for both are given in Table 7.

Table 7. Continuity marker *luego*

	Rueda		Vida		Guía		Suma	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
<i>luego</i> (as DM)	5	0.67	1	0.13	5	0.53	5	0.67
<i>luego</i> (all)	36	5.65	61	7.94	64	6.73	62	8.26

- (28) ANGÉLICA. *Ya que havéis perdonado a la gitana, señor padre, hazed cuenta que las perdonanças son hoy generales.*
 ACARIO. *Así es la verdad.*
 ANGÉLICA. *Luego suplicoos que me perdonéis un pec[c]ado.*
 '[...] In that case I beg you to forgive me a sin' (Rueda, *Medora*, VI)

3.2.5 Closure markers

The closure markers identified in the texts are shown in Table 8:⁸

Table 8. Closure markers

	Rueda		Vida		Guía		Suma	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
<i>últimamente</i>	0	0.00	0	0.00	0	0.00	5	0.67
<i>en fin</i>	0	0.00	4	0.56	0	0.00	1	0.13
<i>al fin</i>	7	1.10	0	0.00	0	0.00	0	0.00
<i>finalmente</i>	3	0.47	0	0.00	26	2.73	6	0.80
<i>por lo demás</i>	0	0.00	0	0.00	1	0.11	0	0.00

7. Angélica takes advantage of her father's forgiving disposition in order to ask forgiveness herself; it is unlikely that *luego* has the meaning of 'immediately' here.

8. Surprisingly, in the light of Cano Aguilar (2007, pp. 13–14), I found no instances of *al cabo* as a DM in these texts.

Vida has overall rather fewer instances of closure markers than the other texts. Between these, there are again differences of lexical preference: *al fin* (as a DM) occurs only in *Rueda*, and while in *Suma* there is variation between the *-mente* adverbs *últimamente* and *finalmente*, in *Guía finalmente* is strongly preferred. There are also other differences of detail. In *Suma*, *finalmente* and *últimamente* appear to be used interchangeably ((29) and (30)), exclusively for introducing a final argument or conclusive summary of a series of arguments, and in initial position, presumably followed by an intonational juncture, as indicated by a comma in the modern editorial punctuation:

- (29) *Últimamente, por concluir y cerrar esta materia, digo que es expresa determinación de la sede apostólica [...] que es usurero el mercader que vende al fiado más caro que al contado [...]*
 ‘Lastly, to conclude and close this subject [...]’ (the last in a series of arguments concerning credit and cash prices) (*Suma*, II, xiv)
- (30) *Finalmente, no hay usura que no se justifique [...] si se admite esta falsedad que es de menor valor la acción y facultad de cobrar de una buena dita que la cantidad.*
 ‘Finally, there is no lending for interest which is not justifiable if [...]’ (*Suma*, II, xvii)

In *Guía*, on the other hand, we seem to have continuing evidence of the extension of *finalmente* to a DM function. While there are many instances of *finalmente* used in much the same way as (30) above (e.g., (31)), there are also some, like example (32), which could be read simply as temporal adverbs and hence not as DMs at all:

- (31) [...] *pregúntalo a san Agustín, a san Ambrosio, a san Gregorio, y finalmente a todos los santos [...]*
 ‘Ask Saint Augustine, Saint Ambrose, Saint Gregory, and indeed all the saints’ (*Guía*, I, viii)
- (32) *Por mí, señor, naciste en un establo, por mí fuiste reclinado en un pesebre, por mí fuiste circuncidado al octavo día, por mí fuiste desterrado en Egipto, y por mí finalmente perseguido y maltratado con infinitas maneras de deshonras e ignominias.*
 ‘[...] and, lastly, you were persecuted and ill-treated [or: you were persecuted and ill-treated at the end] with infinite kinds of dishonour and shamefulness’ (*Guía*, I, vi)

Where a listing consists of clauses, *finalmente* may introduce a logical rather than a temporal finality, that is to say, a conclusion or final overriding argument:

- (33) *Verías la grandeza destes tesoros, verías cuán ciegos andan todos los amadores deste siglo pues no buscan este bien, y verías finalmente con cuánta razón nos convida el Salvador a esta manera de vida [...]*
 ‘[...] and then you would see how right the Saviour is in inviting us to this way of life’
 (Guía, I, vii)

In this way, the stage evidenced in *Suma* is reached, with examples that are very similar to the uses of *finalmente* and *últimamente* found in that text: (34), from *Guía*, is comparable to (29) and (30), from *Suma*:

- (34) *Finalmente, por mucho hablar pierde el religioso el vigor de su ánimo e inquieta a sí y a los otros.*
 ‘Finally, through speaking a great deal, a religious man loses the strength of his spirit and disquiets himself and others’
 (Guía, Tratado enviado al P. Fray Jerónimo de Ferrara)

In summary, the frequency of closure markers in itself is, as with other order markers, probably attributable to text type, since both *Suma* and *Guía* exhibit listings of arguments or points with the marking of an end term; but the difference in specific lexical choice of closure marker in these texts would seem to depend on authorial preference.

3.3 Consecutive connectors

3.3.1 ‘pues’

In its value as a consecutive DM, sentence-initial *pues* indicates a logical conclusion and is broadly commutable with *de modo que* or *por consiguiente*. *Guía* shows a strikingly high frequency of *pues* in this function, while *Rueda* has no such instances. This finding is consistent with Cano Aguilar’s (2007, p. 37) observation that the use of *pues* as a consecutive connector (the equivalent of modern *así pues*) declines in the 16th century but is still attested in ‘textos argumentativos, doctrinales o científicos’. While it may, therefore, have been a marker of such genres, the reason for its maintenance in *Guía* may be that many examples appear to be associated with the syntactic patternings of rhetorical argumentative structure which are characteristic of the sermon: *pues* refers to a precondition for a question (35) or a command (36), which is sometimes also more explicitly expressed by a factive conditional protasis (37).

- (35) *Pues, ¿quién no será llevado por estas cuerdas, quién no se dejará prender destas cadenas, quien no será vencido con tantos beneficios?*
 ‘So [therefore], who would not be drawn by these cords [...]?’
 (Guía, I, vi)

- (36) *Pues no os mueva persuasión de algún hombre a lo contrario desta regla que yo os he dado.*
 ‘So [therefore] let no man persuade you to the contrary [...]’
 (Guía, Tratado enviado al P. Fray Jerónimo de Ferrara)
- (37) *Pues ya, si un hombre pretende alcanzar y conservar siempre aquella soberana virtud de la devoción [...] ¿cómo será posible alcanzar y conservar este afecto tan sobrenatural y tan delicado, si se descuida en la guarda de sí mismo?*
 ‘So then, if [indeed] a man strives to reach and maintain for ever that sovereign virtue of devotion [...]’
 (Guía, I, xiv)

Note that these usages are different in nature from what structurally appear to be similar contexts in Rueda, where, as we have seen, a question following the (commentary) *pues* is generally (see Section 3.1 above) a request for further information rather than the logical conclusion of an argument (6) (another example is given in (38)):

- (38) PANARIZO. *¿Qué remedio había de tener, sino dexar la espada?*
 HONZIGERA. *¿El espada?*
 PANARIZO. *El espada.*
 HONZIGERA. *Pues, ¿el espada havías de dexar sabiendo a lo que vamos?*
 ‘[...] So [let me get this straight], you left your sword [as security] even though you knew where we’re going?’
 (Rueda, *Jauja*)

Pues as a consecutive DM occurs in all texts sentence-internally, where it may be supposed that it is marked off by intonational junctures (39); while *Guía* again shows the highest frequency, there are not inconsiderable numbers of such usages in Rueda, and we may therefore conclude that this use of the consecutive DM is not register-specific. The sentence-initial consecutive *pues* may also be followed by an intonational juncture (40),⁹ but this is clearly impossible to reconstruct with confidence, and modern editors understandably indicate this erratically (cf. example (41), in which no comma has been inserted). The occurrence of both types is shown in Table 9.

- (39) *Dígoos de verdad que antes pasará el cielo y la tierra que una jota o un punto de la Ley se deje de cumplir, hasta que todas las cosas sean hechas. Cualquiera, pues, que quebrantare uno de estos mandamientos pequeños y así lo enseñare a los hombres, pequeño será llamado en el reino de los cielos.*
 ‘[...] Anyone, then, who breaks one of these small commandments and so teaches [other] men, will be called small in the kingdom of heaven’
 (Guía, II, v)

9. This is again consistent with the observations of Cano Aguilar (2007, p. 38).

Table 9. *pues* as a consecutive marker

	Rueda		Vida		Guía		Suma	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
Initial	0	0.00	19	2.47	125	13.14	13	1.73
Internal	13	2.04	5	0.65	62	6.52	8	1.07
Total	13	2.04	24	3.12	187	19.66	21	2.80

- (40) *Pues, para el buen gobierno desta parte, conviene que todas las veces que habláremos, tengamos atención a cuatro cosas [...]*
 ‘Therefore, for sound control of this, we should pay attention to four things every time we speak’ (Guía, I, x)
- (41) [...] *así también se debe aconsejar a los flacos en esta parte, avisándoles que no miren por entonces a sí ni a sus pecados pasados. Pues dirás: ¿En qué debo mirar para cobrar esa manera de esfuerzo y confianza?*
 ‘[...] So you will say, [...]’ (Guía, I, xii)

3.3.2 *así*?

Like *pues*, *así* is multivalent and its various usages are sometimes difficult to distinguish.¹⁰ I have counted those cases where it has a consecutive meaning denoting ‘in this way’, and presents the preceding material as a reason for or cause of what follows, as in (42). It may appear initially, or conjoined to the preceding sentence by *y*,¹¹ and is sometimes, but not always, followed by an indication of juncture. I also indicate in Table 10 instances of what may perhaps be taken as a clefted version of this function, *es así que...*, which is restricted to *Vida* (43), and the few cases of *así* used together with *pues* (44) (not, incidentally, computed into the figures for *pues* in Table 9),¹² as well as the conjunction *así que* (45) with which *así* used as a DM often appears to be commutable.

- (42) *Después vendió a sus mofadores como se le antojaba, porque él solo tenía aceite. Así, en espacio de ocho meses ganó gran suma de dinero [...]*
 ‘[...] In this way, in the space of eight months he earned a large sum of money’ (Suma, II, xx)

10. On the value of *así* in Medieval Castilian see Narbona Jiménez (1978, 281ff).

11. Cano Aguilar (2007, p. 34) observes that the presence of *y* is almost obligatory when it introduces a logical conclusion argued by the speaker.

12. Cano Aguilar (2007, p. 37 n.15) says that *así pues* is sporadically attested as a DM towards the end of the 16th century: the examples from my texts suggest an earlier date.

Table 10. *así* as a consecutive marker

	Rueda		Vida		Guía		Suma	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
<i>así</i>	10	1.57	126	16.41	52	5.47	106	14.13
<i>es así que</i>	0	0.00	15	1.95	0	0.00	0	0.00
<i>así pues</i>	2	0.31	0	0.00	2	0.21	1	0.13
<i>así que</i>	10	1.57	29	3.78	14	1.47	18	2.40

- (43) *Querría escarmentasen en mí los padres para mirar mucho en esto. Y es así que de tal manera me mudó esta conversación, que de natural y alma virtuoso no me dejó casi ninguna [...]*
 '[...] So it was that this conversation changed me in such a way that it left almost no trace of my soul's natural virtuousness' (Vida, 2.4)
- (44) *Cierto es que el usurero prestando corre riesgo y también quien fía ropa, y ni el uno ni el otro puede interesar por ello. Así pues, vender a tanto más del justo precio es delito; no lo abona correr riesgo.*
 '[...] In this way, therefore, selling at so much above the fair price is a sin [...]' (Suma, II, xvi)
- (45) *En esta parte no nos deja de dar la fe que profesamos, ni en la otra la experiencia de lo que cada día vemos. Así que no puede nadie excusar este trago, que sea emperador que sea papa.*
 '[...] So no one, emperor or pope, can excuse this experience' (Guía, I, ii)

As can be seen from a comparison of Tables 9 and 10, the texts differ significantly with regard to the frequency and functions of *así*. *Vida* and *Suma* have a significantly higher frequency of *así* than of *pues* as an initial consecutive DM. *Rueda* always uses *así*, never *pues*, as an initial consecutive DM (could this be because of the preponderant usage of *pues* in *Rueda* as a commentary DM?), while *Guía*, on the other hand, shows a preference for *pues* and a relatively low, even though substantial, attestation of consecutive *así*; it is possible that this is connected with the strong attestation of *así* in other functions in *Guía*, especially the marking of an analogical relationship and its use in comparative formulae (46) (not, of course, counted as DMs), which are shown in Table 11:

Table 11. *así* in analogical and comparative functions

	Rueda		Vida		Guía		Suma	
	raw	ptt	raw	ptt	raw	ptt	raw	ptt
<i>así</i> expressing an analogical relationship	0	0.00	17	2.21	80	8.41	12	1.60
<i>así</i> in comparative formulae	13	2.04	5	0.65	45	4.73	37	4.93

- (46) [...] *para que así como el Señor os perdonó, así vosotros perdonéis* [...] 'So that **just as** the Lord forgave you, **so** you should forgive' (Guía, II, vi)

4. Conclusions

The differences that have been revealed among the four texts studied have generally been differences in the presence, absence or frequency of DMs in a particular function. This strongly suggests that attestation of DMs is dependent on the text type examined, and that the nature of the text determines the DMs used. This accounts for the higher frequency of some kinds of ordering markers (Sections 3.2.1, 3.2.4 and 3.2.5) observed in the didactic and academic texts, where they serve to identify and hierarchise points made in connection with an argument. Such correlations between DM and text type may be seen, then, as consequent primarily upon the tenor of the text. Some other phenomena seem to be more likely to do with the field of the text: the frequency of contrastive spatial order markers (Section 3.2.3) in *Vida*, for instance, may be connected with Teresa's repeated drawing of contrasts and making of distinctions, which is also a feature of *Guía*.

Some other differences, however, are not quite so obviously accountable for in these terms. We have seen how *pues* used in a commentary function is predominant in Rueda (Section 3.1), while in *Guía* its main value is as a more exact consecutive marker (Section 3.3.1). The phenomenon appears to be connected with characteristically different discourse environments: the use of commentary *pues* in the dialogue of Rueda typically to introduce a change of turn (and in *Vida* as a general introductory marker of coherence), while in *Guía* consecutive *pues* is used to express a parenthetical consequence. It seems that in dialogue and in the more 'spontaneous' style of Santa Teresa the strict logical relationship expressed by consecutive *pues* has been 'bleached' to become that of a general coherence marker (cf. Martínez García, 1990, p. 600), who characterises the function of sentence-initial *pues* as 'mero recurso fónico destinado a mantener el «hilo» de la conversación o a dotar de una mejor expresividad al discurso').

We have noticed one or two instances where it does seem possible to speak of variation in the exponents of the same functional category rather than simply the presence or absence of a category. This may amount to nothing more than lexical variation, the data representing a point in the historical trajectory of the rise and fall of near-synonyms: *primeramente* ~ *lo primero* ~ *lo uno* (Section 3.2.1), *mayormente* ~ *especialmente* (Section 3.2.2) and *últimamente* ~ *finalmente* (Section 3.2.5). Some of these variations may be due to what may loosely be termed 'style', since they seem not to be motivated by the nature of the texts concerned,

nor to be correlatable with social class or register. Others, however, may be social markers or indicators (more likely the former in the case of Rueda, where they may be the result of the author's deliberate decision to exaggerate characteristics of the speech of a social class). There may also be evidence of a structural shift in the use of *así* as a consecutive marker in Rueda and *Guía* (Section 3.3.2), occasioned, respectively, by the weakening of *pues* in a consecutive function in Rueda and by the use of *así* in other functions in *Guía*.

Thus evidence has been presented that the dimensions of text type, register, social class and style are of some significance in the description of discourse markers in 16th-century Spanish, and therefore, it may be supposed by extension, throughout the history of the language. A variationist perspective must therefore form the basis for investigation of changes in this area.

References

- Bosque, I. (1989). *Las categorías gramaticales: Relaciones y diferencias*. Madrid: Síntesis.
- Cano Aguilar, R. (2005). La sintaxis del diálogo en el *Quijote* (1615). *Boletín de la Real Academia Española*, 85, 133–155.
- Cano Aguilar, R. (2007). Conectores de discurso en el español del siglo XVI. *Lexis*, 31, 5–45.
- Eberenz, R. (1994). Enlaces conjuntivos y adjuntos de sentido aditivo del español preclásico: *Otrosí, eso mismo, asimismo, demás, también, aun*, etc. *Iberorromania*, 39, 1–20.
- Fernández Alcaide, M. (2009). *Cartas particulares en Indias del siglo XVI. Edición y estudio discursivo*. Madrid & Frankfurt: Iberoamericana & Vervuert.
- Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31, 931–952.
doi:10.1016/S0378-2166(98)00101-5
- Garcés Gómez, M. del P. (2008). La representación de los marcadores discursivos en un diccionario histórico. Propuestas metodológicas. In M. del P. Garcés Gómez (Ed.), *Diccionario histórico: nuevas perspectivas lingüísticas* (pp. 203–234). Madrid & Frankfurt: Iberoamericana & Vervuert.
- González Manzano, M. (2018). *Realmente, verdaderamente* and *ciertamente*: On the relation between epistemic modality and intensifying adverbs. In M. Bouzouita, I. Sitaridou & E. Pato (Eds.), *Studies in historical Ibero-Romance morpho-syntax* (pp. 275–301). Amsterdam: John Benjamins. (in this volume) doi:10.1075/iHLL.16.12gon
- Iglesias Recuero, S. (2000). La evolución histórica de *pues* como marcador discursivo hasta el siglo XV. *Boletín de la Real Academia Española*, 80, 209–305.
- Mancho Duque, M. J. (2008). 'El Camino de Perfección': génesis y aspectos lingüísticos. Alicante: Biblioteca Virtual Miguel de Cervantes. Retrieved from <<http://www.cervantesvirtual.com/obra/el-camino-de-perfeccion-gnesis-y-aspectos-lingsticos-0>>
- Martín Zorraquino, M. A., & Portolés, J. (1999). Los marcadores del discurso. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 3; pp. 4051–4215). Madrid: Espasa Calpe.
- Martínez García, H. (1990). Del *pues* «temporal» al causal y contrastivo. In *Actas del Congreso de la Sociedad Española de Lingüística, XX Aniversario, Tenerife, del 2 al 6 de abril de 1990* (Vol. 2; pp. 599–610). Madrid: Gredos.

- Narbona Jiménez, A. (1978). *Las proposiciones consecutivas en español medieval*. Granada: Universidad de Granada, Secretariado de Publicaciones.
- Real Academia Española/Asociación de Academias de la Lengua Española (2009). *Nueva gramática de la lengua española*. Madrid: Espasa.
- Pons Rodríguez, L. (2010). Los marcadores del discurso en la historia del español. In Ó. Loureda Lamas & E. Acín Villa (Eds.), *Los estudios sobre marcadores del discurso en español, hoy* (pp. 523–616). Madrid: Arco/Libros.
- Pountain, C. J. (2009). Variation in Address Forms in 16th-Century Spanish Prose Drama. In S. Reinheimer-Ripeanu (Ed.), *Studia Lingvistica in honorem Mariae Manoliu* (pp. 282–293). Bucharest: Editura Universităţii din Bucureşti.
- Pountain, C. J. (2012). Dislocación popular y dislocación culta en la comedia en prosa del Siglo de Oro español. In B. Wehr & F. Nicolosi (Eds.), *Pragmatique historique et syntaxe/ Historische Pragmatik und Syntax. Actes de la section du même nom du XXXI^e Romanistentag allemand/Akten der gleichnamigen Sektion des XXXI. Deutschen Romanistentags (Bonn, 27.9. – 1.10.2009)* (pp. 140–156). Frankfurt: Peter Lang.
- Sánchez Jiménez, S. U. (2008). Marcadores discursivos en el teatro de Lope de Rueda. In C. Company Company & J. G. Moreno de Alba (Eds.), *Actas del VII Congreso Internacional de Historia de la Lengua Española. Mérida (Yucatán), 4–8 septiembre de 2006* (Vol. 2; pp. 2163–2188). Madrid: Arco/Libros.

Corpus

- Corpus del español* = Davies, M. (2002-). *Corpus del Español: 100 million words, 1200s–1900s*. Retrieved from <<http://www.corpusdelespanol.org>>
- Rueda = Dialogic material from the following: Hermenegildo, A. (2001). *Lope de Rueda: las cuatro comedias*, edition. Madrid: Cátedra; González Ollé, F. & Tusón, V. (1999). *Lope de Rueda: Pasos*, edition Madrid: Cátedra; and other texts from *Teatro Español del Siglo de Oro*. Madrid: Chadwyck-Healey.
- Vida* = Santa Teresa de Jesús, (1994 [1562]). *Libro de la vida*, edición de Dámaso Chicharro, 10th edition. Madrid: Cátedra.
- Guía* = Fray Luis de Granada, (1567 [1556]). *Guía de pecadores* Morgan Editores, (n.d.).
- Suma* = Tomás de Mercado, (1977 [1571]). *Suma de tratos y contratos*, edición digital basada en la edición de Madrid, Instituto de Estudios Fiscales, Ministerio de Economía y Hacienda.

Studying Ibero-Romance before 1200

Roger Wright

University of Liverpool

Those who wish to study the semantic, morphological or syntactic development of some aspect of Old Spanish need to include in their investigations data from before 1200. The history of Latin in the Iberian Peninsula, before the so-called “Twelfth-century Renaissance”, was similar to that of any other literate living language; most texts (including legal texts) were in the written mode of the native language, despite aiming for traditional orthography, and thus they can be used, with care, as material for the study of the spoken language of their time.

Keywords: Latin; Early Romance; Early Ibero-Romance; *latinorromance*; scribes

1. Introduction

The purpose of this chapter is to suggest that those who wish to study the semantic, morphological or syntactic development of some aspect of Old Spanish need to include in their field of investigations data from a chronologically earlier time than previous scholars have usually done.

I have recently examined a couple of doctoral theses on the development of specific aspects of Old Spanish syntax. One of them concerned the indefinite article which derived from Latin *unus*, as used in Spanish in both the singular and the plural, and the other considered the gradual loss of participial concordance in perfect tenses which use the auxiliary *haber*. Both were good theses which deservedly gained the Ph.D., and I would prefer to leave their authors anonymous, but the authors can be reassured that these theses are only being mentioned here as examples of a wider and almost universal tendency among those who study such developments in this chronological period. Despite their excellence, both theses left me feeling uneasy, and it took a while to realise that it was for the same reason in each case. These analyses, as many others have, investigated copious data from the earliest written Ibero-Romance texts, as they should, but nothing from the

immediately preceding period; and they both also referred, not in great detail, to standard Latin as the starting point of the evolution in question, with quotations and examples taken from the time of the Roman Empire, and no later. The unsettling implication of this procedure seems to be that there is no relevant evidence to be adduced from texts of the intervening thousand years.

This concentration on the earliest written Romance and the Latin of the Empire, but not the intervening centuries, has often in the past been the standard procedural assumption of handbooks on the history of the Spanish language, and indeed of the history of all Romance languages. There seems to have been a general idea that the analyses of the evolutions in question ought to start with the standardised Latin recommended by Roman grammarians, and then jump straight on to the earliest attested texts in unmistakably reformed Romance writing, probably on the grounds that any description of everything that happened in between is effectively guesswork. It would be a mistake to exaggerate this tendency, given the general acknowledgement of the brilliance of Menéndez Pidal's *Orígenes del español* and most Hispanic linguists' awareness of the existence of the Glosses of San Millán and Silos, but evidence of this kind is not what is being referred to; the point is that there is a huge amount of potentially relevant data available for the period in between, not just the eccentrically written words in those Glosses and in some of Menéndez Pidal's documents.

2. Early Romance

James O'Donnell, one of the best scholarly investigators of what we now call Medieval Latin, has made the point (on the LT-ANTIQU discussion site) that, until the 8th century, at least, every century after the so-called end of the Roman Empire has left more written Latin material surviving for us to look at than survives from every century during the Roman Empire. As regards the earliest Medieval centuries, the fact that much of the Latin textual evidence is unappetisingly ecclesiastical does not prevent this merely quantitative fact from holding true. The mathematics include texts from where Latin is a foreign language, such as Germanic and Celtic speaking areas, but it also applies even if we confine our attention to the countries where the native speech of everybody is a kind of Late Latin, or, if we prefer to look at the question from a Romance perspective, as Hispanists usually do, Early Romance. It is also true that some of the Early Romance-speaking writers of the sixth and seventh centuries were deliberately imitating models of the past; not so much because they were using formularies – apart from anything else, Alice Rio's recent book (2009) has demonstrated that these supposed "formularies" are mainly a kind of modern mirage –, but more often because the writers

concerned looked at older texts and tried to write the same way. But on the whole even the most sesquipedalian writers of the so-called Visigothic period (which has almost nothing to do with Visigothic culture, as is now generally realized), before the invasion of 711, were writing in a high style of the language which they actually spoke, and they called both the language they spoke and the language they wrote *lingua latina*. As is normal everywhere, their speech and their writing were seen as two modes of the one language. This means that even 7th century Hispanic scholars with old-fashioned linguistic interests and antiquarian instincts can attest evolved Early Romance features in their writings, such as most noticeably St Isidore of Seville, because that is what they spoke; it is probably common knowledge now, even among Latinists, that some of Isidore's proposed etymologies require an evolved Romance pronunciation to work, such as, for example, the presence of a prothetic initial vowel before words beginning with an [s] plus a consonant (e.g., Nicolas, 2012).

There are more of these Romance features in texts of that time than initially meet the eye, because Early Romance usages do not necessarily have to be different from the Latin ones; that is, there are a large number of features of traditional Latin which are also features of Romance, including a huge amount of vocabulary. For example, if Isidore uses a word such as *tanto* with exactly the same meaning as it had had six centuries before, which is also exactly the same meaning as it was still going to have six centuries later, he does so because it is a 7th century Early Romance usage. The fact that this word has hardly, or not at all, changed meaning or use over the centuries does not stop it being a part of 7th century Early Romance just as much as those words which have. Corominas had not understood this point, as is pointed out in my chapter in the Menéndez Pidal commemorative volume (Wright, 2010). That is, in his *Etymological Dictionary* (1954), Corominas gave the first attestations of *tanto* as being found in much later texts than Isidore. In fact, it appears much earlier than Isidore, with the same spelling and the same meaning, in Plautus. But even when Isidore uses a word which was not used in the Latin of the Empire, Corominas does not want to accept that Isidore's might be the first attested Ibero-Romance use; e.g., the word *tío*, 'uncle', written by Isidore as *thius*, and borrowed from Greek, possibly even by Isidore himself. Corominas writes several erudite paragraphs about this, but still does not allow it to be the first attestation of the Ibero-Romance word, which is said to happen much later; that is, he seems to conflate Romance speech with the arrival of reformed spelling, which is a confusion worth disentangling. The point is that individual Early Romance features chronologically preceded the spelling reforms that later made Romance visible on the page by a long way. There are many such lexical examples.

This point also applies to semantic developments. I pointed out thirty years ago, with a number of detailed 7th century examples, that those who are studying

the semantic development of a word between the Romans and Alfonso el Sabio, should not omit to consider the evidence of the writers of the Visigothic period, including their copious ecclesiastical texts (a study reprinted in Wright, 1995a). I was even then echoing a complaint made by Bill Rothwell (1981), to the effect that if we make hypotheses concerning the semantic development of a word, we should at least at the same time make an effort to see if the textual attestations of novel uses of that word correspond chronologically, if only approximately, to the semantic stages we are wishing to hypothesise.

3. Early Ibero-Romance

The 8th century is short on documentary evidence in the Iberian Peninsula, although not in the Romance area in general, as Marieke Van Acker at Ghent University in Belgium in particular has been demonstrating with her analyses of the language of Merovingian hagiographic texts of the age (e.g., Van Acker, 2007). Then from the 9th century on, texts written by Ibero-Romance speakers begin to get more abundant; that is, texts written by native speakers of the language which modern students of Early Ibero-Romance are ostensibly investigating. This comment even applies to data from the south; it is true that the linguistic analysis of Latin texts from 9th century Moslem Córdoba is tricky, but their awkwardness does not mean that they should all be ignored, as I have shown in studies of the Acts of the Council of Córdoba in 839 (Wright, 2005, 2006). As we move into the 10th century, we reach, in the Northern part of the Peninsula, the earliest manifestations of the copious documentation which Menéndez Pidal exploited so brilliantly many decades ago. That is, we have reached what has come to be called the *Época de orígenes*, a label inspired by Menéndez Pidal's rather unfortunate book title of 1926, referring essentially to the tenth and eleventh centuries. But large amounts of evidence from legal and ecclesiastical sources of that epoch have been published in the years since 1926, particularly in León, dating from the 9th century onwards, and there is no sensible excuse for ignoring it and starting historical linguistic analyses from the years after 1200. The point is not that these sources from the *Época de orígenes* contain words which are spelt in peculiar ways, although they often do; words spelt in their traditional way, by a scribe who we can tell for that very reason was attentive in the initial spelling lessons of his *formación*, are no less able to attest the Romance semantics and morphosyntax of the scribe's native speech.

The study of semantic change in this period, although rarely carried out in this way, has been becoming increasingly easier in the sense that data can now come to hand more quickly, as we are able to look at attestations of individual

words in relevant computer-assisted concordances. Furthermore, it is important to make the point that the same goes for the study of syntactic and morphological change in those years, mainly because so many of the developments of Romance concerning morphosyntax involve what are in effect semantic changes in particular words. The complaint which is sometimes heard that syntactic data cannot be found in lexical repertoires, and that every sentence has to be tagged in some way with its syntactic features before analysis can be possible, is not entirely accurate. When I was studying, in 1993, the use of reflexive syntax without reflexive semantics in the period between the Romans and the 13th century – that is, the use of the word *se* as an indicator of passive meaning without necessarily implying an agent, rather than of reflexive meaning where the same entity is both agent and patient (Wright, 1995b) – the first step was to look for uses of the word *se* in relevant texts. In 1993 it was difficult to be anything like exhaustive in such searches, naturally, and at that time we did not yet have available the large concordances of the Latin documentation there is in the Archives of León Cathedral; when those concordances appeared, it became possible to work towards a more nuanced analysis than before. In 2002, I wanted to study both the semantic and the syntactic evolution of *sedere* in León during the so-called *Época de orígenes*. Latin *sedere*, meaning ‘to sit’, is the origin of Spanish *ser*, meaning ‘to be’, having thus undergone a semantic change, and it came to be regularly used as a passive auxiliary, which was a morphosyntactic change. By 2002, in order to find relevant data, there was no need to do anything more complicated than look in those concordances from León. Before the mid-13th century the texts which are now in those Archives were prepared in unreformed spelling. However, the writers were speakers of Ibero-Romance, and until the 12th-century Latin reforms, which really kicked in during the second half of the century within the Kingdom of León, the writers were naturally using words with their normal contemporary meanings, whatever the spelling was. The results were presented as being a partial study of the semantic development of *sedere* (Wright, 2004), but, in effect, it was a syntactic study as well, of the word’s use as an auxiliary for the passive. If the analysis had confined itself to 13th-century usage in the new Romance spellings, operating on the assumption that previous data in unreformed spellings could not be relevant, and if as a consequence 13th-century Ibero-Romance *ser* had been directly contrasted only with traditional Latin *sedere* as attested in Imperial times, there would have been less to say.

The study of the history of *ser* and *estar* exemplifies this theme with remarkable neatness. There is a study by José María Saussol of the uses of *ser* and *estar* in the *Poema de Mio Cid*. It is entitled “*Ser*” y “*estar*”: *orígenes de sus funciones en el Cantar de Mio Cid* (Saussol, 1977). The book’s title, using the word *orígenes*, suggests that the first attestations of the newly developed usages of these verbs cannot have

been earlier than the *Poema de Mio Cid*. This is unfortunate. Regardless of what date is given to the *Poema de Mio Cid*, no historical linguist would want to imply that nobody spoke any kind of Romance before that poem was composed, and that, at that point, Ibero-Romance suddenly turned up overnight, fully formed, even in the mouths of those who had never come across the *Poema*. Saussol never explained what he thought was happening with these verbs in speech in the years immediately before their first appearance in the new written Romance forms. It is actually rather a good book, once we accept that Saussol is investigating these developments only during a period many centuries later than their actual origins, *origenes*. For in startling contrast there is also a brilliant book by Birte Stengaard, Professor of Spanish in the University of Oslo, called *Vida y muerte de un campo semántico. Un estudio de la evolución semántica de los verbos latinos stare, sedere e iacere del latín al romance del s. XIII* (Stengaard, 1991). That is, Birte Stengaard ends her study of the development of these verbs at the chronological point where José María Saussol starts his; Birte Stengaard, as is the case with Scandinavian scholars in general, but not many others, is able to trace the history of these verbs and their meanings in the Iberian Peninsula through all the intervening years from the Roman Empire up to Alfonso el Sabio, with many relevant examples and acute analysis. Scholars working in Scandinavia understand this possibility, maybe because of their admirable tradition of close analysis of Medieval Latin texts from Romance-speaking areas; for a further example, Timo Riiho, Professor of Ibero-Romance languages at the University of Helsinki, had much the same chronological canvas in view in his lengthy study of the semantic and syntactic development of the prepositions *por* and *para*, entitled “*Por*” y “*para*”: *estudio sobre los origenes y la evolución de una oposición prepositiva iberorrománica* (Riiho, 1979). For Birte Stengaard and Timo Riiho, at least, the *Época de Orígenes* begins with the Romans and extends up to the 13th century. But the Scandinavians, though absolutely right, are unfortunately exceptional, and Menéndez Pidal’s use of the word *origenes* to refer to nothing earlier than the latish 9th century, though much preferable to Saussol’s perspective that the origins of Spanish started at the time of the *Poema de Mio Cid*, has led to its own problems.

As regards the two theses mentioned at the start, it would not have cost much effort to look at uses of *unus* and *habeo* in the León Concordances, at least. If there happened not to be any, that would be a most interesting discovery too, for converse reasons. For as well as pointing to Romance manifestations in texts in unreformed Latin orthography, it is sometimes worth wondering why a few of the features which have been postulated by modern Romanists as being common in Vulgar Latin appear only rarely in written texts of that period. This applies to *se*, in fact, which is less common in Medieval Latin texts, in any meaning, than we would expect in the light of the fact that the word is used all over the place when

written Romance at last turns up undisguised. In that study of the agentless use of *se* (Wright, 1995b), the other constructions were also considered which could be used instead for the same purpose, that is, for representing agentless passive semantics, including the old-fashioned synthetic passive morphology and the intransitive use of verbs which were usually transitive. Although it is not easy to look for uses of synthetic passive morphology via a lexical concordance, following up this train of thought more recently, I prepared a study (Wright, 2013a) concerning the ostensibly increasing use of deponent verbs in writing during this period, an increase which worries Romanists inasmuch as it happened at a time when Romanists would prefer to think the deponent category was moribund. The conclusion of that investigation is that scribes were encouraged in their training to use the obsolescent passive morphological inflexions in writing as much as they could, but often without their really understanding the semantic purposes which the forms had once been used for. This is the reason behind why they look like deponents when in fact they are normal actives with a *-ur* ending added in writing in order to look proper. In the same way, it seems clear that the scribes were discouraged in their training from using as many prepositions in their writings as they used in their speech. It is not difficult to think of close parallels in the teaching of the writing of English nowadays, where common phenomena of speech are discouraged by teachers from appearing in writing.

I was already turning over in my mind the implications of these trains of thought, when an invitation arrived to review the recent *Lexicon of Medieval Latin from the Kingdom of León (Lexicon Latinitatis Medii Aevi Regni Legionis (s. VIII-1230) Imperfectum)*, edited by Maurilio Pérez González (2010), Professor of Latin at León. The data on which it is based cover the whole of León, rather than just the city, but strictly Asturias and León, rather than Galicia or Castile (whose similarly conceived lexica are being prepared separately). I was hoping to be able to tell those readers of the review who are interested in the study of Ibero-Romance before 1200, that this is the source of Early Ibero-Romance information which we have been waiting for. And potentially it is, as far as strictly lexical studies are concerned; all words qualify to be included, since, as Maurilio Pérez says, almost every Latin word has undergone some kind of semantic development over the centuries, even *et*. But they are not all in this *Lexicon* yet. Unfortunately, the *imperfectum* of the title means ‘unfinished’; and words of direct syntactic interest are not included in this initial volume. They will be there in due course, a course currently envisaged as becoming due in about the year 2017, but the present state of the *Lexicon* does not include, for example, the prepositions, an exhaustive study of which would be needed, and could be used, for the analysis of the evolution of sentence structure over those centuries; and also missing so far are other words whose semantic changes are

central to the study of morphosyntactic developments, such as the words *habere*, *unus*, *sedere* and *se*, which have been considered in this present contribution, and also *ille*, *sum* and others.

The heart of this Lexicon is in the right place, though, for it has acquired its data from the forty-five documentary collections from León which had been published by the year 2001, plus all the historiography of the Astur-Leonese realm. Its hinterland is correspondingly large, and the result is of huge value. ‘Medieval Latin’, as used in the Lexicon’s title, means anything written in the relevant geographical area after the African Invasion of 711 and before the union of León and Castile in 1230. Maurilio Pérez is admirably aware that calling these texts ‘Latin’ at all might be a bit problematical, so he has used a word to refer to their language which I had occasionally used previously when writing in Spanish: the word *latinorromance*, written as one compound word with no hyphen. In explaining why he uses this word he tells us that he sees the language of the time as still being ‘Latin’, which is the case, but the Latin written by Romance speakers of the age, and thus much influenced by their Romance. Whereas in contrast I see the language of these texts as being the other way round, as being the Romance of the age much influenced by the linguistic details which the scribes had had instilled into them in their training (e.g., Wright, 2013b). This latter scenario, whereby the Ibero-Romance-speaking scribes learnt a number of tricks of the trade in their training for use when transferring their speech into the written mode, and thus made the resulting text look more like Latin, seems more likely than the alternative scenario, of thousands of scribes laboriously learning Latin as a complete separate system and then going to great lengths to get it wrong, which seems to be the general view of the disciples of Rafael Lapesa. Maurilio Pérez and I got into an (unpublished) discussion about these two perspectives at the Fourth Hispanic Medieval Latin conference in Lisbon a few years ago, but our views are close. Several other factors suggest that it might be sensible to regard these texts as in essence Romance rather than Latin; for example, the presence in texts of the age of a number of contemporary words of Arabic origin, many of which would have been used naturally in normal spoken Romance by the 12th century but had never appeared in old traditional written Latin, seems a possible indicator of the fact that usually the scribes were wanting to represent their own contemporary language on paper rather than to write a different language belonging to the distant past. Relevant attestations of these words of Arabic origin are included here in this first part of the Lexicon (for a review, see Wright, 2013c).

Meanwhile, Maurilio Pérez himself seems dangerously interesting, or at best off-message, to some of his Latinist colleagues in Spain, who still view any text whose writer was aiming at traditional orthography as being thereby an example of Latin; “bad” Latin, “corrupt” Latin, “decadent” Latin, “barbarous” Latin,

maybe, but still Latin, rather than Romance. In Portugal, António Emiliano has led a splendid counterattack which goes further than Maurilio Pérez or even I do (e.g., Emiliano, 2005), but this ebullient rationality is still not normal in Spain. The late Manuel Ariza's recent book on Spanish in the 12th century (Ariza, 2009) explicitly avoids these questions by the simple expedient of looking only at specifically reformed written Romance data, and then saying that he disagrees with me on these matters, without explaining to his readers either what he thinks instead, or, indeed, what I think. Ariza's book is actually rather a charming work in its concentration on detailed data allied to an insouciant avoidance of anything theoretically interesting, thereby leaving a number of loose ends. He even avoids any discussion of the *Poema de Mio Cid* and its language by pointing out that the manuscript was written later than the 12th century, which is uncontroversial (as opposed to the dating of the text). What Maurilio Pérez refers to as *latinorromance*, and I refer to as Early Romance written by scribes attempting to write in the ordinary traditional way, Manuel Ariza calls "latín más o menos macarrónico", which seems to imply a scribal ability to distinguish clearly between Latin and Romance in the 12th century, an ability which his data do not seem to support.

But maybe, slowly, the perspective which seems natural to the Scandinavian scholars such as Timo Riiho and Birte Stengaard, and to me, and to some others, might come in due course to seem natural to the Spanish and other Hispanists investigating the history of Spanish before 1200. Essentially, we can see the history of Latin, before the advent of the purist reforms connected with the so-called 12th-century Renaissance, as being similar to that of any other literate language. Until the written forms of such a language are deliberately updated to reflect its evolved nature, the continued training of all scribes to use the older traditional prescribed forms comes more and more to lead to the written orthographical form being a disguise for the spoken form of the language rather than a representation of it. The earliest experiments at spelling a language in a new way, such as the Riojan Glosses, look odd in retrospect, in the same way that text messages do nowadays, but, like text messages, they were new ways of representing the same language, not a new language. Then after a while some enterprising souls with the authority to do so decided to undertake a radical overhaul of the whole written mode, and invent written Romance. In due course, since the old written mode survived as well, people came to think of them as being two languages rather than two ways of writing the same language. But the texts that survive from the age before these new experiments were undertaken, although they overlaid the speech of the scribes with something of a disguise, are still potential repositories of many interesting data of value to our studies of the history of Ibero-Romance, and cannot just be ignored.

In conclusion, there is no more reason to start the study of Ibero-Romance phenomena with the first texts prepared in a new written form in about 1200 than with those of any other date. Indeed, I have said most of this before, but it does need to keep on being said.

References

- Ariza, M. (2009). *La lengua del siglo XII (dialectos centrales)*. Madrid: Arco/Libros.
- Corominas, J. (1954). *Diccionario crítico etimológico de la lengua castellana*. Madrid: Gredos.
- Emiliano, A. (2005). Representational models vs. operational models of literacy in Latin-Romance legal documents (with special reference to Latin-Portuguese texts). In R. Wright, & P. Ricketts (Eds.), *Studies on Ibero-Romance linguistics dedicated to Ralph Penny* (pp. 17–57). Newark, DE: Juan de la Cuesta.
- Fernández Catón, J. M. (Dir.) (1987–2010). *Colección documental del Archivo de la Catedral de León (años 775–2000)*. León: Centro de Estudios e Investigación «San Isidoro», 26 Vols.
- Menéndez Pidal, R. (1926). *Orígenes del español. Estado lingüístico de la Península Ibérica hasta el siglo XI*. Madrid: Espasa-Calpe.
- Nicolas, Ch. (2012). Etymologizing from eye to ear: about vowel prosthesis in Isidore's etymologies. In F. Biville, M. -K. Lhommé & D. Vallat (Eds.), *Latin vulgaire – latin tardif IX* (pp. 795–806). Lyon: Publications de la Maison de l'Orient et de la Méditerranée.
- Pérez González, M. (Ed.) (2010). *Lexicon Latinitatis Medii Aevi Regni Legionis (s. VIII–1230) Imperfectum/Léxico latinorromance del reino de León (s. VIII–1230)*. Turnhout: Brepols.
- Riiho, T. (1979). “Por” y “para”: *Estudio sobre los orígenes y la evolución de una oposición prepositiva iberorrománica*. Helsinki: Societas Scientiarum Fennica.
- Rio, A. (2009). *Legal practice and the written word in the Early Middle Ages: Frankish formulae, c. 500–1000*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511581359
- Rothwell, W. (1981). Historical semantics and the structure of Medieval French vocabulary. In T. E. Hope, T. B. W. Reid, R. Harris, & G. Price (Eds.), *Language, meaning and style: Essays in memory of Stephen Ullmann* (pp. 145–155). Leeds: University of Leeds Press.
- Saussol, J. M. (1977). “Ser” y “estar”: *Orígenes de sus funciones en el Cantar de Mio Cid*. Sevilla: Publicaciones de la Universidad de Sevilla.
- Stengaard, B. (1991). *Vida y muerte de un campo semántico. Un estudio de la evolución semántica de los verbos latinos stare, sedere e iacere del latín al romance del s. XIII*. Tübingen: Max Niemeyer.
- Van Acker, M. (2007). *Ut quique rustici et inlitterati hec audierint intellegant: Hagiographie et communication verticale au temps des mérovingiens (VIIe–VIIIe siècles)*. Turnhout: Brepols.
- Wright, R. (1995a). The study of semantic change in Late Latin (Early Romance). In R. Wright (Ed.), *Early Ibero-Romance: Twenty-one studies on language and texts from the Iberian Peninsula between the Roman Empire and the thirteenth century*. (pp. 65–73). Newark, DE: Juan de la Cuesta.
- Wright, R. (1995b). La sintaxis reflexiva con semántica no agentiva. In M. Pérez González (Ed.), *Actas del I Congreso Nacional de Latín Medieval (León 1–4 de diciembre de 1993)* (pp. 415–431). León: Universidad de León.

- Wright, R. (2004). La representación escrita del romance en el Reino de León entre 1157 y 1230. In J. M. Fernández Catón (Ed.), *Orígenes de las lenguas romances en el reino de León, siglos IX-XII* (Vol. 1; pp. 273–293). León: Centro de Estudios e Investigación «San Isidoro».
- Wright, R. (2005). Nominal morphology in the Acts of the Council of Córdoba (839): A sociophilological investigation. In R. Archer, V. Astvaldsson, S. Boyd & M. Thompson (Eds.), *Antes y después del Quijote. En el cincuentenario de la Asociación de Hispanistas de Gran Bretaña e Irlanda* (pp. 497–505). Valencia: Biblioteca Valenciana.
- Wright, R. (2006). La representación de las vocales en las Actas del Concilio de Córdoba de 839. Una investigación sociofilológica. In C. Arias Abellán (Ed.), *Latin vulgaire – latin tardif VII* (pp. 565–572). Sevilla: Universidad de Sevilla.
- Wright, R. (2010). Ramón Menéndez Pidal and the history of the Spanish language. In J. C. Conde (Ed.), *Ramón Menéndez Pidal after 40 years: a reassessment* (pp. 145–162). London: Department of Hispanic Studies, Queen Mary, University of London.
- Wright, R. (2013a). Passive morphology in Late Latin. In A. Garcea, M. -K. Lhommé & D. Vallat (Eds.), *Polyphonia Romana: Hommages à Frédérique Biville* (Vol. 1; pp. 81–90). Hildesheim: Georg Olms.
- Wright, R. (2013b). How scribes wrote Ibero-Romance before written Romance was invented. In E. -M. Wagner, B. Outhwaite, & B. Beinhoff (Eds.), *Scribes as agents of language change* (pp. 71–83). Berlin: De Gruyter Mouton. doi:10.1515/9781614510543.71
- Wright, R. (2013c). Review of *Lexicon Latinitatis Medii Aevi Regni Legionis (s. VIII–1230) Imperfectum*. Edited by Maurilio Perez Gonzalez. Turnhout: Brepols. 2010. pp. lxxvii + 805. ISBN: 978250353473. *Hispanic Research Journal*, 14(2), 194–195.

Index

A

adverb 9, 50, 109, 153, 157, 159,
162, 165, 209, 224, 229–232,
254, 265, 277, 279–287,
289–294, 296, 298–299
adverbial clause 8, 104, 174,
176, 179, 180, 181, 184, 185, 187
adversative 9, 223, 225, 228,
230, 234, 236, 244
agent 127, 129, 135, 138, 173–174,
178, 181, 190, 329
al fin 316–317
allative 8–9, 195–196, 198,
208–210, 217
allomorphy 2–3, 13–18, 20–24,
26–30, 32–41, 43, 45
anthroponym 85, 87
article 2–4, 6, 9, 13, 17, 41,
47, 49, 75–95, 97, 131, 146,
149–150, 181–182, 198, 206,
269, 306, 325
así 10, 84, 92, 242, 285–286,
306, 309, 315, 318–323
asimismo 90, 315, 323
Asturian 1, 3, 13, 25–26, 33,
37, 39
aunque 9, 223–227, 229–235,
244–245, 253, 310, 314

B

basic word order 183–185
Biblical Hebrew 257, 267, 269,
271
Biblical translation 258–260,
266, 269
binomial 47
bleaching 79, 190, 288

C

case-marking 173
Catalan 1, 9, 15–17, 22, 30–31,
34–35, 41, 43, 53, 58, 100, 170,
174, 193, 223–227, 229–231,
235, 237, 239–241, 244–245,
248, 252–253, 265–266, 272

causal 9, 34, 223, 225–230,
237–241, 243–245, 307–308
chunking 8–9, 50, 182, 195, 198,
200–201, 205, 217
ciertamente 9, 275, 277–289,
291–299, 323
clash 32, 36
Classical Greek 111
Classical Portuguese 7, 150,
183–184
clitic 8, 102, 114, 116, 118–119,
186–187, 190–191, 270
clitic pronoun 187, 191
closure marker 318
compound 4, 47–52, 54–55,
57–58, 60–68, 71–73, 110, 112,
180, 332
concatenative 2–3, 47–49,
51–71
concatenative compound
62–64
concessive 2, 9, 180, 223–241,
243–245, 253
subordinator 2, 223, 253, 272
conditional 5, 9, 14, 16, 104,
113, 142, 223, 225–230,
232–233, 237, 239–242,
244–245, 249, 251, 253–258,
265, 267–270, 318
conditional concessive 9,
223, 230, 232–233, 239–241,
244–245, 253
conditional structure 255
conjunction 9, 49, 60, 63,
65–67, 70–71, 109, 133, 197,
223–224, 226, 230–233,
235–236, 238, 240, 244–245,
249, 251, 254, 256–257,
264–266, 269, 307, 320
consecutive connector 308, 318
construction 2–4, 6, 82–83, 89,
101, 103, 106–107, 124, 126,
130–131, 135–136, 138–144,
147, 156, 162, 173–174, 179,
181–182, 186–192, 196, 198,
206–207, 212–213, 217, 229,
239–241, 244, 276, 305, 311
continuity marker 315–316
contrastive focus 8, 155, 159,
181, 184–185, 187–191
coreferential 176, 177, 178, 184,
188, 191
coreferentiality 176–177
corpora 177, 265, 277–278,
303–304
corpus 7, 60, 74, 78, 89, 92,
97–98, 121, 125, 128, 141,
147, 155, 157, 162, 165–167,
170–171, 177, 180, 183, 194,
196–197, 201–202, 207, 220,
244, 247–248, 250, 263, 273,
278–279, 299, 301, 304–306,
313, 324
cyclic 223

D

dative 5–6, 123–126, 129–132,
134–136, 138–147, 173, 206
dative experiencer 6, 123–125,
129–130, 132, 134–136,
139–140, 142–144
dative subject 6, 130, 147
definite 2, 4, 75–85, 87–95, 128,
131, 135, 146, 181–182, 198,
206, 297
definite article 2, 4, 75–85,
87–95, 131, 146, 181–182, 198,
206
demás de esto 312
disambiguation 8, 184, 188,
189, 191, 192
discourse function 183–185,
189–190
discourse marker 10, 276, 281,
284, 286, 289–291, 293, 299

E

empathy 139–140
emphasis 75, 79, 91, 93, 105,
107, 111, 131, 138, 187–189, 238

- en fin* 316
en parte 314
encara que 9, 223–227,
 229–230, 235–236, 244, 253
 ending 19, 26–28, 32, 34–35,
 60, 113, 276, 331
 entrenchment 173, 191
 epistemic 9, 179, 263–264,
 275–277, 280, 283–284, 286,
 289–290, 293–296, 298–299,
 303
 epistemic adverb 286, 296, 298
 epistemic modality 275, 277,
 280, 283
 epistemicity 275, 277, 286
especialmente 10, 312–313, 322
 exceptional case-marking 173
- F**
finalmente 316–318, 322
 focus 8, 59, 99–100, 105–106,
 111–112, 118, 120, 136, 138,
 150, 154–156, 159, 162–163,
 168, 170, 180–181, 184–185,
 187–191, 223, 228, 239–240,
 252, 278, 289
 Fray Luis de Granada 10–11,
 304
 frequency-based 173
 future 2, 5, 12, 14, 16, 49, 113,
 261, 311
- G**
 Golden Age Spanish 190
 grammaticalisation 2, 5,
 8–10, 78–79, 95, 119, 168,
 195, 209, 217, 223, 231–232,
 234, 237, 241, 243–244, 249,
 251, 256, 258, 267–270, 275,
 277–278, 288–289, 293–294,
 298–299, 303
gustar 2, 6, 123–136, 138–145,
 147
- I**
 Ibero-Romance 1–3, 8, 11,
 13–14, 17–18, 20–21, 24, 28,
 30, 32, 36, 38–39, 41, 43, 55,
 71, 73, 119, 146–147, 170, 174,
 193–194, 219, 246, 269, 272,
 300, 303, 323, 325, 327–334
 indicative 2, 13, 15, 17, 20,
 22, 26–28, 31, 34–37, 226,
 232–236, 239, 241–245,
 254–255
 Indicator 161, 303, 329, 332
 inergative 149–150, 153, 158,
 160, 166, 168
 infinitival 7–8, 173–191
 infinitival clause 173, 178, 181,
 183, 186, 188
 infinitive 7–8, 14, 16, 99,
 112–113, 129, 173–174,
 176–186, 188–193, 195–197,
 199, 208–217
 inflected infinitive 173,
 177–178, 184, 192
 information structure 6–7,
 149–150, 154, 156–157,
 160–161, 169, 183, 192
 insubordination 5, 249,
 268–269
 intensification 181, 188,
 191–192, 277
 intensifier 8, 180–181, 191
 intensifying 10, 181, 275, 277,
 289–290, 292–299
 intensifying adverb 289,
 292–294
 interrogative structure 249,
 251–254, 257, 259, 263, 269
 intersubjectification 4, 75, 78,
 85, 95, 97
ítem 311–312
- L**
 Latin 1, 3–5, 11, 18–19, 21,
 25–27, 30–31, 39–40,
 43–44, 49, 51–55, 67, 72,
 75, 78–80, 94–97, 100, 108,
 111–112, 119, 126–127, 130,
 132, 135, 145–146, 195, 200,
 203, 218–219, 224, 231,
 249–251, 253–259, 262–264,
 266–267, 269–273, 276,
 279, 282, 300, 325–327,
 329–335
 linking vowel 3, 47–49, 54–55,
 59, 62–63, 65–71
lo primero 10–11, 310–311, 322
lo segundo 310–311
lo uno 310–311, 322
 Lope de Rueda 10, 303–304,
 308, 311, 324
luego 76, 87–88, 91, 113, 135, 137,
 204, 312, 316
- M**
mayormente 10, 312–313, 322
 Medieval Portuguese 182
 Medieval Romance 5, 249,
 254, 265, 269
 Medieval Spanish 6, 67, 74,
 203, 250, 259, 269, 279
 metaphor 127, 144, 146
 modality 145, 256, 275,
 277–280, 283, 287, 300
 Modern Portuguese 7, 41, 169,
 178, 180, 184–185, 188
 Modern Spanish 2–3, 14, 16,
 20, 25, 27, 29, 36, 49, 106, 125,
 179, 191, 213, 250, 259, 269,
 305, 307, 309, 313
 morphology 2, 16, 18, 42–44,
 47, 72–73, 142, 157, 178, 184,
 276, 295, 331
 morpheme 2, 13–14, 17, 21, 23,
 28–33, 43
- N**
 new information 6–7, 149–150,
 152–154, 156–158, 162–163,
 165–169, 183, 185
 nominative subject 124–126,
 130, 136, 179
 non-coreferential 176–178
- O**
 oblique 5, 120, 123–125, 143,
 145, 147, 173
 Old Portuguese 7, 39–40,
 149–150, 183, 185
 Old Romance 110, 117
 Old Spanish 2–5, 8, 15–21,
 24–26, 31, 33, 36–37, 39,
 41, 99–102, 104–106,
 108–109, 111–114, 116–118,
 179, 204, 224, 227, 237–238,
 249, 251–252, 259–264,
 267–268, 325
 opening 10, 155–156, 310
 ordering 10, 111, 143, 306, 310,
 322
otrosí 315, 323
 overt agent 173–174
 overt subject 7–8, 102–103,
 173–178, 181, 184–185, 189,
 191–192
 Overt Subject Infinitive (OSI)
 construction

- P**
- para* 2, 8, 50, 73, 76, 84, 88,
90, 92, 145, 147, 156, 159,
169, 171, 175–178, 180–183,
186–187, 194–198, 200–221,
280, 295, 310–312, 320–322,
330, 334
- per més que* 229, 237, 239, 245
- per molt que* 238
- ponderative 237
- por ejemplo* 305–306
- por lo demás* 316
- por más que* 9, 229, 237, 239
- por mucho que* 9, 239–240, 245
- por otra* 10, 307, 314–315
- por otra parte* 307, 314–315
- por una parte* 10, 314
- Portuguese 1–2, 6–8,
13–14, 21, 25–26, 39–41, 44,
53, 58, 135, 149–150, 152–155,
157, 160, 163, 165–166,
168–171, 173–186, 188–189,
191–192, 195, 252–254,
265–266, 269
- postverbal 7–8, 149, 152–153,
156–162, 166–169, 174,
178–185, 187–190, 192, 238,
253, 277, 284–285
- postverbal subject 8, 152, 168,
180, 184, 190
- pragmatic 4–6, 9–10, 75–77,
83, 86, 95, 107, 123, 126, 140,
145, 173, 181, 184, 188–192,
225–226, 229, 276–277,
287–289, 293–294, 299,
301, 308
- pragmatic bleaching 190
- pragmatic cost 225–226, 229
- pragmatic function 77, 181,
191–192, 276
- pragmaticalisation 9, 276,
288–289, 298
- preposition 8–9, 50, 127,
129–130, 145, 173, 180–182,
186, 195–199, 201–208,
210–212, 217, 238–239
- prepositional 115–116, 173, 180,
183, 185–187, 193, 207, 215,
217, 230
- primeramente* 10–11, 310–311,
322
- primero* 10–11, 102, 296,
310–311, 322
- pronominal subject 7, 160,
179–180, 182–190
- proper name 4, 75–87, 89,
92–95, 146
- pues* 10, 94, 104, 234, 261, 264,
307–310, 314, 318–323
- puesto que* 9, 223, 225, 230, 237,
240–245
- purposive 8–9, 195–196, 209,
211–218
- Q**
- Q-particle 224, 249–254,
256–257, 260, 264, 267–270
- quirky 2, 5, 123–124, 143, 146
- R**
- realmente* 9, 275, 277–289,
291–299, 323
- reanalysis 5, 145, 166, 189, 239,
256, 268, 293
- register 10, 305, 309, 319, 323
- retention 217
- rhetorical questions 250, 252,
260
- root 3, 13, 17–18, 21–28, 30–41,
54, 59, 203
- S**
- Santa Teresa 10, 304–305, 322
- scalar 9–10, 230–233, 239, 244,
296–298
- semantic 5–6, 8, 11, 14, 16, 22,
47–50, 54, 57, 61, 67, 70–71,
79, 97, 105, 123–124, 126–128,
131, 134, 136–138, 140, 142,
145–147, 159, 165, 173, 184,
195, 200, 205, 207, 209,
214, 217, 223, 225, 237, 253,
277–278, 286–288, 293–294,
299–301, 325, 327–331
- semantic change 6, 97, 209,
286–288, 293, 301, 328–329
- retention 217
- sense of taste 126–127, 142
- sentence adverb 277, 283–287,
290–291
- si* 2, 5, 76–77, 88, 90, 103–105,
113–114, 116, 134, 142, 153, 198,
202, 223–224, 227–231, 234,
241, 246, 249–251, 253–272,
283, 285–286, 295, 306, 309,
311, 314–315, 317, 319
- sociolinguistic 2, 220, 304
- Spanish 1–11, 13–34, 36–37,
39, 41–45, 47–49, 53, 55–62,
67–68, 70, 73–75, 77–81, 85–
86, 93, 95, 99–102, 104–109,
111–114, 116–120, 123–127,
130–135, 139, 142–145, 160,
173–180, 182, 185, 188–193,
195, 199–200, 202–206,
209, 211–213, 217, 223–227,
229–232, 235, 237–241,
244–245, 247, 249–254,
259–265, 267–270, 275–279,
282, 295, 299, 303–305,
307, 309, 313, 323, 325–326,
329–330, 332–333
- spatial 8, 49, 55, 196, 198, 200,
202, 204, 207–210, 217–218,
313–314, 322
- specified 81, 83, 86, 91, 95, 129,
174, 178, 193
- split 8, 99, 111–114, 118, 120,
142, 185, 187
- stem marker 47
- Structural scope 9, 275,
281–283, 285–286, 288–292
- stylistic fronting 2, 4, 99, 107,
114, 120, 146
- subject 4–8, 18, 77, 89, 99,
102–103, 108–109, 120,
123–127, 129–131, 133, 135–136,
138, 143, 147, 149, 152, 154–155,
157–160, 162–169, 173–192,
206, 211, 214–216, 223, 240,
253, 258–260, 317
- subject position 6–7, 102–103,
108, 143, 154, 159, 166,
168–169, 179–180, 184,
186–187, 190–191
- subjectivity 97, 127
- subjunctive 2, 13, 15, 28, 31,
36–37, 142, 213, 226–227,
232–233, 235–236, 241–244,
254–255
- subordination 177, 223, 268
- SVO 7–8, 108, 150, 152, 154, 157,
183–185, 188
- T**
- text types 78, 177, 299, 303
- textual emphasis 75, 93, 146
- theme 42, 50, 81–82, 228, 329
- Tobler-Mussafia Law 110–111

Tomás de Mercado 10–11, 304
 topic 3, 6, 23, 34, 108, 110–111,
 114, 124, 126, 137, 139, 152, 154,
 156, 162, 166, 183–184, 209,
 223, 230, 315
 topicalisation 99, 101, 159
 topicality 108, 137, 139
 transitive 6–7, 123, 125, 127,
 132–133, 135, 145, 149–150, 153,
 157–163, 165–168, 331
 TVX 108–109, 111, 113–114, 119
 últimamente 10, 316–318, 322

U

unaccusative 149–150, 152,
 154–155, 159–160, 163–169

unaccusative verb 154, 169
 usage frequency 177

V

V2 108, 157–158
 variation 3–4, 6, 8, 11, 13, 15, 17,
 24, 28, 31–34, 37–38, 41, 64,
 67, 77, 95, 100, 124, 157, 170,
 192, 195–196, 198, 202–203,
 205, 210–212, 217–220, 252–
 253, 263, 271, 278, 303–304,
 306, 310, 317, 322
 variationist 2, 120, 214, 216, 323
 velar allomorphy 2–3, 13–18,
 20–24, 26–28, 33–34, 36–39,
 41, 45

verbigracia 306
 verbs of ‘liking’ 126
 verbs of ‘wanting’ 134
verdaderamente 9, 275,
 277–291, 293–296, 298–299,
 323

W

word order 5, 7–8, 99–101,
 108–109, 143, 149–150,
 152, 154, 157, 159, 162, 166,
 168–169, 173, 183–185, 190,
 192, 253, 259
 WordSmith Tools 306

This volume features fourteen papers by leading specialists on various aspects of historical morpho-syntax in the Ibero-Romance languages. In these papers, fine-grained analyses are developed to capture the richness of undiscussed or – often – previously unknown data. Comparative across the (Ibero-)Romance languages and diverse in terms of the approaches considered, ranging from cognitive-functional to generativist to variationist, they combine in this volume to showcase the merits of different, yet complementary, perspectives in understanding linguistic variation and language change. The gamut of phenomena scrutinised varies from morpho-phonological puzzles and word-formation to syntax and interface-related phenomena to, as a coda, methodological suggestions for future research in old Ibero-Romance; thus making it ideal reading for scholars and postgraduate students alike.

ISBN 978 90 272 0045 7



9 789027 200457

John Benjamins Publishing Company