# Si hubiese un cambio en la morfología del imperfecto del subjuntivo, ¿cómo hubiera ocurrido?: A diachronic analysis of Spanish *-ra*/*-se* morphological variation

by

Hannah M. Harker

B.S., Malone University, 2013

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF ARTS

Department of Modern Languages
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2019

Approved by:

Major Professor
Dr. Mary T. Copple

# Copyright

© Hannah M. Harker 2019.

# Abstract

Modern Spanish has two morphological paradigms at its disposal in the Imperfect Subjunctive, collectively represented by *-ra* and *-se.* The distribution of each morpheme has been much examined, both synchronically (Williams, 1982; DeMello; 1993; Kempas, 2011; Sorenson, 2016; Guzmán Naranjo, 2017) and diachronically (Martínez, 2001; Ramírez Luengo, 2001; Branza & van Heuven, 2005; Hanna, 2012), but questions still remain about how the use of these two forms changed over time. While synchronic studies conducted over the last 30 years have revealed that *-ra* is unequivocally the dominant form in all varieties of Spanish, diachronic analyses have shown that this was not always the case. Data collected from the 15th and subsequent centuries have shown a steady increase in the usage of *-ra* at the expense of *-se* over time. In attempts to trace its path, these studies have used corpus data to primarily report the relative frequency of each morpheme in the various syntactic contexts where they occur. Although a clear picture does not emerge upon comparing these findings, the general tendency for *-ra* to appear more frequently in conditional sentences (i.e. *si tuviera…, podría…* 'if I had..., I could…´) and in other adverbial clauses has been documented (Martínez, 2001; Ramírez Luengo, 2001; Branza & van Heuven, 2005). In view of the need to go beyond relative frequency alone to explain *-ra*/*-se* variation, the present study employs variable-rule analysis in an attempt to more clearly hypothesize the nature of the generalization of *-ra* as it progressed as a grammaticalizing Imperfect Subjunctive morpheme by identifying the factors conditioning the selection of *-ra* over time. Data taken from a corpus of dramas and novels written in Spain in the 17th and 19th centuries were coded for syntactic context (conditional protasis, adverbial, adjectival, nominal clauses), verb frequency, verb polarity and use in dialogue or narration. Findings show that in the 17th century dialogue, conditional protases and other adverbial clauses,

high-frequency verbs, and negative contexts favored the use of *-ra*. In the 19th century, adjectival clauses surpassed protases and adverbial clauses as the most favorable for *-ra*, and dialogue and high-frequency verbs remained favorable while polarity was no longer selected as significant. The key findings of this study suggest that the degree of syntactic embeddedness of clauses (Hopper & Closs Traugott, 2003) influences the ability of an innovative form to spread to new contexts and that routinization of high-frequency forms may help lead the way.

# Table of Contents

List of Figures ..... viii
List of Tables ..... ix
Acknowledgements ..... x
Dedication ..... xii
Chapter 1 - Introduction ..... 1
1.0 Overview of *-ra*/*-se* Variation ..... 2
Chapter 2 - The Grammaticalization of *-ra* and *-se* ..... 7
2.0- Fundamentals of Grammaticalization ..... 7
2.1- The Latin Roots of *-ra* and *-se* ..... 8
2.1.1 The Trajectory of *-se* ..... 9
2.1.2 The Trajectory of *-ra* ..... 11
2.2 The Debated Allomorphy of *-ra* and *-se* ..... 17
2.3 Other Observations of *-ra*/*-se* Distribution ..... 23
2.4 Conclusion ..... 24
Chapter 3 - Analyzing *-ra*/*-se* Variation ..... 25
3.0 Making Sense of Linguistic Variation ..... 25
3.1 Diachronic Variationist Analysis ..... 26
3.2 Corpus Construction ..... 28
3.2.1 Selection of Corpus Texts ..... 28
3.2.1.1 17th Century Texts ..... 29
3.2.1.2 19th Century Texts ..... 30
3.2.2 Extraction ..... 32
Chapter 4 - Coding of Factor Groups ..... 33
4.0 Introduction ..... 33
4.1 Textual Factors ..... 33
4.2 Linguistic Factors ..... 34
4.2.1 Verb-related Factors ..... 34
4.2.2 Clause-related Factors ..... 34
4.3 Functions of the Imperfect Subjunctive ..... 40

4.3.1 Temporal Reference. 40
4.3.2 Degree of Realization According to Context 41
4.4 Selection of Factors for Analysis 43
4.4.1 Excluded Factor Groups 43
Chapter 5 - Constraints on *-ra*/*-se* in the 17th and 19th Centuries 45
5.0 Introduction 45
5.1 Relative Frequency of *-ra* and *-se* 45
5.2 Independent Clauses and the Generalization of *-ra* 46
5.3 Dependent Clauses and the Generalization of *-ra* 47
5.4 17th Century Multivariate Analysis 50
5.4.1 Dialogue/Narration 51
5.4.2 Syntactic Context 53
5.4.2.1 Degree of Clause Dependence and Grammaticalization 53
5.4.3 Verb Frequency 56
5.4.4 Polarity 59
5.5 19th Century Multivariate Analysis 61
5.5.1 Pathways of Change 63
Chapter 6 - Conclusions and Directions for Future Research 66
6.0 Summary of Findings 66
6.1 Limitations and Directions for Future Research 67
6.2 Concluding Thoughts 68
Bibliography 69
Appendix A - Codes for In-text Examples 74
17th Century 74
18th Century 74
19th Century 74
20th Century 74
Appendix B - Descriptions of Corpus Texts 75
17th Century Dramas 75
La vengadora de las mujeres 'The Avenger of Women' 75
Amor, honor, poder 'Love, Honor, Might' 75

El amor constante ‘Constant Love’ ..... 76
17th Century Novels ..... 77
La gitanilla ‘The Gypsy Girl’ ..... 77
El buscón ´The Swindler´ or Historia de la vida del Buscón, llamado Don Pablos, ejemplo de vagamundos y espejo de tacaños ´History of the Life of the Swindler, Called Don Pablos, Model of Vagabonds and Mirror of the Stingy´ (Books I and II) ..... 78
19th Century Dramatic Texts ..... 78
El sí de las niñas ´The Maiden´s Consent´ ..... 78
Don Álvaro o la fuerza del sino ´Don Álvaro or the Force of Destiny´ ..... 79
El afán de figurar ´The Thirst for Appearances´ ..... 80
19th Century Novels ..... 80
El Senor de Bembibre ´The Gentleman from Bembibre´ (Chapters 1-10) ..... 80
Escenas andaluzas ´Scenes from Andalucía´ (Pages 1-48 of the PDF version available online through the Biblioteca Virtual Universal) ..... 81

# List of Figures

Figure 1: The Evolution of *-se* ..... 11
Figure 2: The Evolution of *-ra* ..... 16
Figure 3: Progression of Clause Dependence ..... 40

# List of Tables

Table 1: Relative Frequency of *-ra*/*-se* in the *Corpus de español* ..... 4
Table 2: Relative Frequency of *-ra*/*-se* in Speech ..... 5
Table 3: Relative Frequency of *-ra*/*-se* in Newspapers ..... 5
Table 4: 17th Century Texts ..... 30
Table 5: 19th Century Texts ..... 31
Table 6: Total *-ra*/*-se* Token Counts by Century and Text ..... 32
Table 7: Relative Frequency of *-ra*/*-se* in the 17th and 19th Centuries ..... 45
Table 8: Relative Frequency of *-ra*/*-se* in Dependent Clauses ..... 49
Table 9: Multivariate Analysis Results for the 17th Century ..... 51
Table 10: Distribution of Verb Frequency by Clause Type (17th Century) ..... 58
Table 11: 19th Century Multivariate Analysis Results (with 17th Century for Comparison) ..... 62
Table 12: Distribution of Verb Frequency by Clause Type (19th Century) ..... 64

# Acknowledgements

First and foremost, I give thanks to God, the Author of life itself, whose grace and strength sustained me throughout the writing of this thesis. I´m thankful for the opportunity to study in this program and feel that I have been extremely blessed by this experience. I will forever cherish the relationships I've built with classmates, professors, and others in Manhattan.

I'd like to express my deepest gratitude to Mary Copple, my Major Professor, for all her nurturing guidance throughout the completion of this project, beginning at its inception in her Advanced Spanish Grammar course. I´m thankful for the amount of time she generously spent to teach me the ropes of multivariate analysis and to help me refine my argument. I´m also incredibly grateful for her supportive mentoring, her willingness to always lend a listening ear, and for the wise advice she´s shared about research, teaching, and life.

I also gratefully acknowledge the support of my committee members, Laura Valentín-Rivera and Rebecca Bender. I thank Dra. Valentín-Rivera for always challenging me to think more deeply and thoroughly and for teaching me the fascinating intricacies of the history of Spanish. I thank Dr. Bender for helping answer numerous questions related to using literature for linguistics research and for showing me that studying and teaching literature can be fun. I also thank Dr. Pablo Martínez Diente for corpus text suggestions. Lastly, I´m grateful for the support I´ve received from other professors in the Department of Modern Languages, who are too numerous to list, over the last few years.

I am also greatly appreciative of the support of my classmates and friends here in Manhattan. I thank Alejandra, Chrissie, Daniel, Daniela, Emili, Joe, Joey, Kelsie, Kwasi, Natacha, and Victoria for inquiring often about my progress and for letting me go on about it. Most of all, I thank my officemate, Logan, especially for the days when he graciously allowed

our office to become a sea of books and journal articles. Furthermore, I thank Irma, Rachel, Julie, Jenny, Alyssa, Hannah, Laurel, and Allison, whose ever-present friendship has been a blessing and much-needed at times.

Last, but certainly not least, I thank my family and friends in Ohio, Wisconsin and elsewhere. I'm beyond grateful for their constant love and support as I've ventured to Kansas to pursue my Masters. They have always been my biggest cheerleaders, and they mean the world to me.

# Dedication

In remembrance of my grandma Alice Mae, who passed away after a valiant fight against cancer during the completion of this project. Words cannot capture how selfless and caring she was nor can they fully express my gratitude for the example she set for my family. Among the many, many things she taught me was the importance of work ethic and how it is not just about completing a task but rather seeing to it that it has been completed well. There is much about her I admire and aspire to be.

# Chapter 1 - Introduction

Languages are embodied by a rich and dynamic system that is constantly evolving through a variety of mechanisms of change (Bybee, Perkins & Pagliuca, 1994). Innovations may appear in any level of the system (e.g. phonological, morphological, syntactic, lexical) and can reflect processes of simplification (the merger of multiple forms to one function) or of diversification (the division of one form into multiple functions). Internal and external factors may act upon the system to produce these innovations. Linguistic change, however, is not contingent upon the introduction of a new form or the inception of a novel use of a preexisting form alone. The innovation must take root and increase in frequency as it competes with already existing forms which serve the same function. The level of sociolinguistic prestige the innovation receives has been suggested as one possible driving force of such acceptance (Labov 1994; 2001 in Pharies, 2015). As innovations are accepted and gain frequency, they can displace the previously existing form, causing it to eventually fall into disuse. This process can be quite slow, often taking centuries, and is composed of many stages of coexistence (Bybee *et al.*, 1994; Penny, 2004). Examining the change in frequency between competing variants over time in their contexts of use (both linguistic and social) can shed light on the nature of the expansion of innovative forms.

One process known to lead to these changes is grammaticalization (Hopper & Closs Traugott, 2003). The study of grammaticalization, which will be described in greater detail in Chapter 2, has developed within the field of language variation and change with the aim of determining how lexical items and expressions lose all or part of their lexical meaning and come to serve grammatical functions. The development of English *be going to* to express futurity as opposed to its strict lexical meaning of movement toward a place (see Hopper & Traugott, 2003,

p. 1) and Spanish perfects formed with the auxiliary verb *haber*, which originally expressed possession (see Pharies, 2015, pp. 3-4), are oft citied examples of this process. Additionally, grammaticalization research also examines the processes by which existing grammatical forms gain new grammatical functions (Bybee *et al.*, 1994; Hopper & Closs Traugott, 2003). The morphological variation that has developed in the Spanish Imperfect Subjunctive (and by extension Pluperfect Subjunctive) can be taken to exemplify this latter definition.

## 1.0 Overview of *-ra*/*-se* Variation

In Modern Spanish, speakers may utilize two morphological paradigms to encode the Imperfect Subjunctive: one represented by *-ra* and the other represented by *-se*[1]. As seen in examples (1) and (2), this variation is not restricted to the Imperfect Subjunctive, but also applies to the Pluperfect Subjunctive, formed by the auxiliary verb *haber* and a past participle.

(1) . . .como si todas esas cosas **fueran** inevitables y él las **hubiera sabido** desde siempre.

'. . .as if all these things **were** inevitable and he **had** always **known** them.'

TÍA 20th cen.[2]

(2) . . .procuró que la suya **fuese** la más excelente de quantas **se hubiesen escrito**.

'. . .[s/he] procured that his/hers **was** the most excellent of all those which **had been written.**'

GRAM 18th cen.

[1] In this paper, language specific grammatical categories will be capitalized (e.g. Spanish Imperfect Subjunctive) while the general term will not (i.e. references to the subjunctive in general). Also, *-ra* and *-se* will be respectively used to represent the full morphological paradigm (e.g. *-ra* = *-ra, -ras, -ra, -ramos, -rais, -ran*).

[2] The texts from which examples have been taken are represented by an abbreviated name along with its century of publication. A list of the source texts along with their codes can be found in Appendix A. All translations are the author's own, except when otherwise indicated as having come from the example source.

Furthermore, as seen in examples (3) and (4), both forms may appear in the same sentence.

(3) **Hubiérasle pedido** a Dios que te **diese** algo más de entendimiento.

'You **would have asked** God that he **should give** you a little more understanding.

SDB 19th cen.

(4) No sé lo que **hubiera dado** porque **hubiese podido** oírla.

'I don't know what you (formal) **would have given** so that you **would have been able** to hear her.'

SDN 19th cen.

Like many features of Modern Spanish, these morphemes were inherited from the Latin verbal system. A more comprehensive description of this process will be given in Chapter 2, but, in essence, *-ra* and *-se* evolved from the Latin Pluperfect Indicative and Pluperfect Subjunctive paradigms respectively. Early in the development of Hispano Romance, *-se* gained its modern value of Imperfect Subjunctive, followed much later by *-ra* (Pitloun, 2006; Penny, 2004). Since entering into variation, however, *-ra* has gradually been replacing *-se* and is unequivocally the most commonly used Imperfect Subjunctive morpheme today. This is confirmed by an exhaustive search of Imperfect Subjunctive tokens (including *haber + participle*) in the *Corpus de español: Historical/Genres*[3] (Davies, 2002) presented in Table 1.

[3] The *Corpus de español: Historical/genres* is an open-access data base of authentic written and oral texts comprised of 100 million words representative of texts from the 12th century to the 20th century. It includes collections of legal documents, religious texts, personal letters, fictional literature, journalism and interviews (20th century only) from all Spanish-speaking countries (pre-16th century texts are assumingly only from Spain).

| **Relative Frequency of *-ra*/*-se* between the 16th and 20th Centuries the *Corpus de español* (Davies, 2002)** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **16th Century** | | **17th Century** | | **18th Century** | | **19th Century** | | **20th Century** | |
| | **N** | **%** | **N** | **%** | **N** | **%** | **N** | **%** | **N** | **%** |
| ***-ra*** | 34,079 | **30** | 31,797 | **49** | 12,965 | **32** | 41,815 | **56** | 32,118 | **90** |
| ***-se*** | 80,225 | **70** | 32,675 | **51** | 28,120 | **68** | 32,797 | **44** | 3,554 | **10** |
| **Total** | 114,304 | **100** | 64,472 | **100** | 41,065 | **100** | 74,612 | **100** | 35,672 | **100** |

**Table 1: Relative Frequency of *-ra*/*-se* in the *Corpus de español***

Moreover, oral and written data collected during the last 30 years from various Spanish-speaking countries have revealed that this change has taken place in all varieties of Spanish. As attested by DeMello (1993) and Sorenson (2016), *-ra* has, by far, become the dominant morpheme in each of the analyzed regions (at best, *-se* composed only 20% of the collected tokens); however, it is important to note that *-se* has not fallen into disuse in any of the examined varieties, neither in speech (DeMello, 1993) nor in writing (Sorenson, 2016)[4]. Furthermore, the average of all cities analyzed in both studies (calculated here from the two studies´ results for comparative purposes) is almost identical, suggesting that genre differences do not significantly influence the selection of one form over the other in contemporary use. Tables 2 and 3 show the cities with the highest and lowest use of *-ra* from each respective study along with the average of all cities analyzed.

[4] DeMello took data from oral corpora from Bogotá, Buenos Aires, Caracas, Havana, Lima, Madrid, Mexico City, San Juan, Santiago, and Sevilla. Sorenson examined newspapers from the capital city of each Spanish-speaking country except for Equatorial Guinea.

| **Distribution of *-ra*/*-se* in Speech Adapted from DeMello (1993)** | | |
|---|---|---|
| | *-ra* | *-se* |
| Lima, Peru | 99% | 1% |
| San Juan, Puerto Rico | 80% | 20% |
| Average (10 cities) | 93% | 7% |

**Table 2: Relative Frequency of *-ra*/*-se* in Speech**

| **Distribution of *-ra*/*-se* in Newspapers Adapted from Sorenson (2016)** | | |
|---|---|---|
| | *-ra* | *-se* |
| Mexico City, Mexico | 97% | 3% |
| Asunción, Paraguay | 79% | 21% |
| Average (20 cities) | 92% | 8% |

**Table 3: Relative Frequency of *-ra*/*-se* in Newspapers**

Time will tell if *-se* will eventually fall into disuse. However, the trend revealed by the *Corpus de español* (Davies, 2002) data does suggest that, as Penny (2004) comments, "the eventual resolution of this morphological variation can be safely predicted" (p. 216).

While variation in these subjunctive forms has been investigated in several studies, the scope of analysis has been narrow. Many investigations have been synchronic, establishing the relative frequency of the two forms in specific regional contexts in a given time period (Williams, 1982; DeMello, 1993; Kempas, 2011; Sorenson, 2016; Guzmán Naranjo, 2017; McKinnon, 2018). The few diachronic studies that have been done (Martínez, 2001; Ramírez Luengo, 2001; Branza y van Heuven, 2005) focused on relatively short periods of time, typically a century or less, and their aim also revolved around presenting the relative frequencies of the two forms in various linguistic and, to a much lesser degree, sociolinguistic contexts. Klein-

Andreu (1991) offers an analysis of Pluperfect Indicative *-ra* that helps pinpoint how it became a subjunctive. However, none have attempted to explain the path of generalization of subjunctive *-ra* as it competed with *-se* over time by statistically analyzing the distribution of the two morphemes to identify those contexts which first exhibited innovative use of *-ra*. Thus, this study aims to build on this previous work to provide an in-depth analysis of *-ra*/*-se* diachronic variation. Before delving into this discussion, however, a brief history of the grammaticalization of the two forms along with more specific findings of the work previously done on this variation will be presented.

# Chapter 2 - The Grammaticalization of *-ra* and *-se*

## 2.0- Fundamentals of Grammaticalization

Cross-linguistically, grammaticalization has proven to be a common mechanism of language change. Through processes of reanalysis, wherein lexical items, such as pronouns, verbs or entire constructions (e.g. *be going to*), are interpreted in novel yet communicatively compatible ways, these forms can take on new meanings and, consequently, be used in new contexts (Hopper & Closs Traugott, 2003). As use of the form in these new contexts becomes more frequent and routinized, grammatical nuances of the form's use may become more salient than its original lexical meaning. Thus, with time, the form can lose its strict lexical meaning to assume an increasingly grammatical function. However, it should be noted that the form may continue to be used with its original meaning alongside its grammaticalized counterpart, a phenomenon known as layering. Additionally, as Bybee *et al.*(1994) demonstrate, grammatical forms, such as subjunctives, may also evolve over time due to similar processes.

Grammaticalizing forms have been found to share three main characteristics: 1) an increase in frequency, 2) generalization of meaning which allows them to appear in additional communicative contexts (e.g. *be going to* extending from its strict sense of movement to future intention), and 3) rigidification of syntax in that these forms tend to become more fixed as a unit (e.g. *gonna* must be followed by a bare infinitive). Similar to this last point, phonological reduction is another common characteristic of grammaticalizing forms. Of relevance for the present study is the increase in frequency of *-ra*, demonstrated above in Table 1, and *-ra*'s presumed spread to new subjunctive contexts of use as it competed with *-se*. The data set utilized in this study was taken from points in time in which *-ra* and *-se* competed as Imperfect

Subjunctive inflections. Nevertheless, to better situate their status as competing morphemes, we will first examine the individual trajectories which led to their variation in the first place.

## 2.1- The Latin Roots of *-ra* and *-se*

The Latin verbal system underwent significant restructuring as Hispano Romance, the precursor of Modern Spanish, took shape between the 5th and 13th centuries AD (Azevedo, 2009). The four Latin conjugation paradigms simplified into three (-ĀRE, - ĒRE, -ERE, - ĪRE > -AR, -ER, -IR); novel verb forms, such as the conditional (e.g. *yo cantaría* ´I would sing´) and morphological future (e.g. *cantaré* ´I will sing´) were added; and many inflectional paradigms, such as those for the passive voice, were replaced by periphrastic structures (e.g. AMĀTUR > AMĀTUS EST > *él es amado* 'he is loved') (Pharies, 2015). The two morphemes of interest in this study were conserved as this new verbal system took shape although they no longer served the same functions as their Latin counterparts. As previously mentioned, Latin *-ra* and *-se* were in no way allomorphs, meaning they served different functions and were not interchangeable as they are today. While they both cooccurred with another affix, -VE or -VI, to express pluperfect tense (used to refer to an event which occurred prior to another past event), they did not share mood (used to convey speaker attitudes or the assertability of a proposition). The affix *-vera* belonged to the assertive indicative mood (CANTA**VERA**M > *yo había cantado* 'I had [indicative] sung') and *-visse* belonged to the non-assertive subjunctive mood (CANTA**VISSE**M > *yo hubiera/hubiese cantado* 'I had [subjunctive] sung'). Although *-ra* and *-se* have come to share grammatical functions in Modern Spanish, the two forms developed independently of each other, as will be shown below.

### 2.1.1 The Trajectory of *-se*

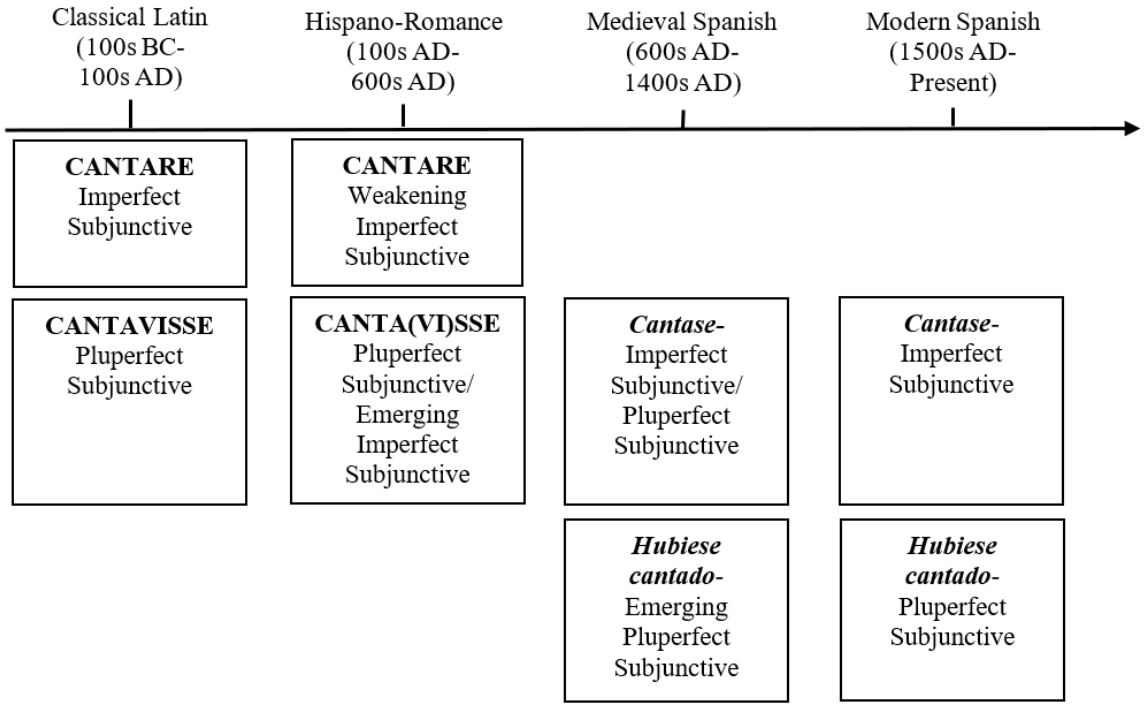
Vulgar Latin, from which the Romance languages were derived, had an imperfect subjunctive morphological paradigm represented by *-re* (CANTA**RE**M > *yo cantara/cantase* 'I sang [imperfect subjunctive]'). It is believed that at the time of the fall of the Roman Empire in the 5th century A.D., the sociopolitical event that facilitated the linguistic diversification of Latin through the disintegration of the once unified empire, the pluperfect subjunctive form with *-se* was already beginning to replace *-re* in imperfect subjunctive contexts (Penny, 2004). Wolford (2005) credits this change to the loss of the perfect grammatical marker *-vi* in Hispano Romance. The reduced form CANTASSEM retained its pluperfect value, but it also came to be increasingly viewed as an imperfect past subjunctive form, analogous to CANTAREM, since it no longer contained an explicit morpheme for perfect value. At the same time, as Ward (1998) suggests, analogical leveling may have further facilitated the loss of the Latin Imperfect Subjunctive encoded by *-re* due to its formal similarity to the developing Future Subjunctive, a merger of the Future Perfect Indicative (CANTAVERO 'I will have sung') and Present Perfect Subjunctive (CANTAVERIM 'I have sung' ). As the emerging Latin Future Subjunctive CANTARE 'I may sing' became more frequent, the strength of *-re* as an imperfect subjunctive form decreased, which enabled *-se* to continue expanding into contexts of use which had previously been exclusively occupied by *-re*.

Penny (2002) points out that imperfect, as opposed to pluperfect, uses of *-se* were first seen in conditional sentences that referred to atemporal hypothetical situations (i.e. *si yo* ***tuviese****..., **pudiese*... 'if I **had**…, I **would be able to**…'). This marked an innovation since the *-se* form had previously only been used in conditional sentences that referred to unrealized or impossible situations of the past (in Modern Spanish *si yo* ***hubiese tenido****..., **hubiese podido*...

'if **I had had**…, I **would have been able to**…'). More specifically, *-se* came to be employed in both the protasis (the conditioning clause) and the apodosis (the conditioned outcome) of these hypothetical conditional sentences (Penny, 2004). Use of *-se* in the apodoses gradually ceded to *-ra*, which in turn ceded to the conditional, encoded by *-ría* (Lapesa, 1981; Klein-Andreu, 1991; Penny, 2002). From the conditional sentence, in particular the protasis, *-se* generalized to other imperfect subjunctive contexts, completely replacing *-re* over time.

The gradual loss of the Latin Imperfect Subjunctive in Hispano Romance was accompanied by the emergence of a new periphrastic pluperfect construction formed by *haber* + *past participle* (with accompanying indicative and subjunctive forms), an innovation which further propelled the gradual erosion of the pluperfect meaning *-se* originally had expressed. The *haber* + *past participle* gained frequency and eventually replaced *-se* as the exclusive Pluperfect Subjunctive form (Pharies, 2015). Like most grammaticalization processes which progress slowly over time (Bybee *et al.*, 1994), this change occurred over several centuries. Lapesa (1981) states that *-se* continued to appear as a pluperfect subjunctive morpheme through the Middle Ages and was not fully replaced by *hubiese* + *past participle* until the 16$^{th}$ century. Figure 1 summarizes the restructuring of Imperfect and Pluperfect Subjunctive morphology from Latin to Modern Spanish (adapted from Wolford [2005] and Pharies [2015]). Due to the fluid nature of language change, the dates below represent general ranges, not definitive start and end points, for the phases of the evolution of *-se*.

**Figure 1: The Evolution of *-se***




### 2.1.2 The Trajectory of *-ra*

When compared with *-se,* subjunctive *-ra* is a much younger and relatively more complex innovation since it involved not only a shift in temporal reference but also in mood. Penny (2002) states that there is evidence to suggest that in as early as late Latin *-ra* had already begun to appear in the apodoses of conditional sentences but did not make the jump to other non-indicative contexts until much later. There is a discrepancy amongst scholars regarding when *-ra* first started to appear in other syntactic contexts with its generalized subjunctive value, with theories ranging from the 12$^{th}$ century to the 14$^{th}$ century (Pharies, 2015; Lapesa, 1981; Penny, 2004); however, it would seem that a later century is more plausible. Klein-Andreu (1991) found that in a 14$^{th}$ century Peninsular literary text *-ra* was still being used more frequently than the

emerging *haber* + *past participle* (N = 135/200, 68%)[5] in indicative pluperfect contexts, underscoring the fact that *-ra* was still strongly competing with *haber* + *past participle* in these contexts (p.169). If *haber* + *past participle* were found to occur at a similarly low rate in other texts of that time, then it could also be asserted that it advanced aggressively to replace Pluperfect Indicative *-ra* over the next century. Indeed, according to Pharies (2015), it is believed that by the 15th century Pluperfect Indicative *-ra* had fallen into disuse, and Lapesa (1981) adds that by the 16th century pluperfect indicative uses of *-ra* were recognized as archaic. The conservation of *-ra*, in light of this, indicates that *-ra* had developed a strong enough association with its new non-indicative functions to withstand being lost as *haber* + *past participle* grammaticalized and, subsequently, replaced it as a marker of pluperfect tense. Pinpointing the exact moment of *-ra*'s initial spread into truly subjunctive territory has perhaps been difficult since, as Veiga (1999) points out, written texts where such changes can be analyzed are not always available, do not necessarily reflect spoken language and are often copies of the original documents made at much later dates, compromising their validity. Nevertheless, the possible pathway *-ra* followed as it evolved into an Imperfect Subjunctive morpheme has been proposed.

Ward (1998) comments that it was potentially nothing less than a "[morpho]syntactic sleight of hand" that gave rise to the subjunctification of *-ra* (p.127). However, Bybee *et al.* (1994) have noted several cases of indicatives developing into subjunctives in other languages when these forms are displaced by new ones fulfilling the same grammatical function, as was the case for *-ra*. Perhaps *-ra*'s transformation should not be considered so surprising if it can be

[5] It should be noted that Klein-Andreu (1991) only analyzed the writing of one work by one author (*El Conde Lucanor* by Don Juan Manuel), so the strong preference for *-ra* in indicative pluperfect contexts may very well reflect the author´s personal preference.

thought of in this way, this is, as a byproduct of the generalization of *haber + participle* as the Pluperfect Indicative form of choice. A fuller discussion of the mechanisms which possibly allowed non-indicative *-ra* to spread into all subjunctive contexts will be taken up later in this section and again in Chapter 5, section 3, but first let us examine how indicative *-ra* was freed of its sole role as an indicative in the first place.

Bybee *et al.* (1994) propose that grammaticalizing forms, such as *haber + past participle* tend to develop their new function in main clause contexts, and then subsequently permeate syntactically subordinate ones. Consequently, when an already existing form, such as *-ra*, is edged out of main clause contexts as the newer gram gains frequency, it tends to be relegated to subordinate contexts, where it may eventually be lost or become available to serve a new function. If no new association is made, then the pre-existing form will most likely be completely replaced by its innovative competitor or highly constrained to specific conservative contexts or collocations. However, if the older gram gains a new function, it will also most likely develop first in main clause contexts since it can no longer serve its original function there once the newer gram has taken over in those asserted contexts. In particular to subjunctives that arise from indicatives, the fact that the older gram with its indicative value gets relegated to subordinate contexts is what consequently allows it to become more associated with non-assertion and thus gain subjunctive value (Bybee *et al.*, 1994). In the same vein, Klein-Andreu (1991) argues that it was indicative *-ra´*s increasingly constrained use as a temporally remote form in pragmatically low-focus subordinate clauses that allowed it to dislodge itself from indicative assertion to be used as an assertive form in main clauses which communicated hypothetical situations. As with *-se*, conditional sentences have been identified as the most probable locus of change for subjunctive *-ra* (Lapesa, 1980; Penny, 2002; Klein-Andreu, 1991).

As previously mentioned, Penny (2002) states that as early as late Latin *-ra* was beginning to be used as a "conditional" form in the apodoses of conditional sentences since the periphrastic pluperfect was replacing it in other contexts. While this use is not considered truly subjunctive, the use of the term "conditional" to separate it functionally from *-se* should be used with caution. The term "conditional" seems to have been assigned in virtue of the fact that non-indicative *-ra* was highly constrained to the apodoses of conditional sentences, a clause in which *-se* had also been used previously (Penny, 2002). As with *-se, -ra* initially appeared with pluperfect temporal reference in these sentences (i.e. *si yo hubiese tenido…,* ***pudiera*** (=***hubiera podido***) 'if I had had…, I **would have been able to**…'). In this way, non-indicative *-ra* was used first to refer to counterfactual hypothetical events of the past[6]. Eventually the modern Conditional, encoded with the *-ría* paradigm, replaced *-ra* in the apodoses of irrealis and counterfactual conditional sentences (Pitloun, 2006). However, it is important to point out that use of *-ra* in the apodosis of conditional sentences remained frequent through the 16th and 17th centuries and can still be observed throughout the Spanish-speaking world today, particularly in Venezuela and Central America (Penny, 2004; Butt & Benjamin, 2011). It is also worth noting the fact that *-ra* was also not lost as Conditional *-ría* replaced it in conditional sentences, providing further evidence that *-ra* had acquired another meaning/function, beyond its residual Pluperfect Indicative or specialized irrealis function, prior to or alongside this competition.

Following its appearance in conditional apodoses, *-ra* spread to protases that expressed impossible or unrealized situations of the past, drawing on its roots in pluperfect temporal

[6] Interestingly, *-ra* was overwhelmingly favored in counterfactual contexts in the data of this study (80% or greater occurrence). Whether this demonstrates retention of this early favored context or the influence of other factors (e.g. syntactic context of the token) has not been analyzed.

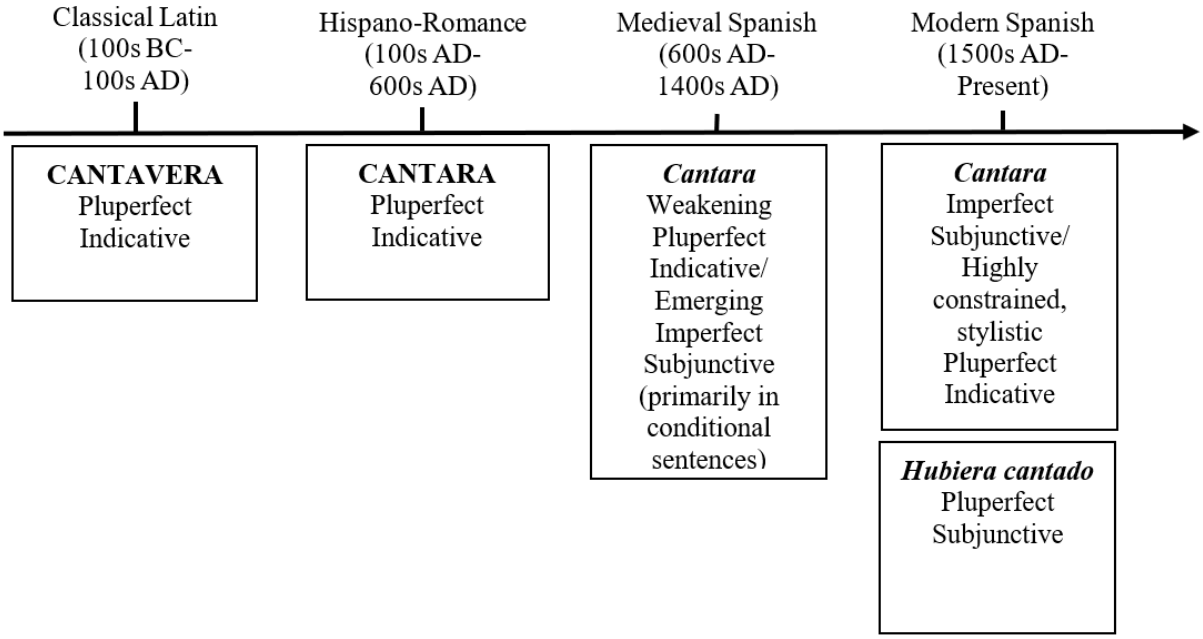
reference. This extension, while innovative, is not entirely unexpected given that it is quite possible for verb forms to assimilate across clauses in conditional sentences (Lavandera, 1983 in Montrul, 2018). For example, in certain varieties of Modern Spanish, the Conditional is sometimes used in both clauses, as shown in (5) adapted from Silva-Corvalán (2001, p.3).

(5) Si **tendría** veinte millones de pesetas, yo no **trabajaría** más.

'If I **would have** twenty million pesetas, I **would** not **work** anymore.'

With time, *-ra* also came to be used in atemporal conditional sentences which expressed improbable hypothetical situations (Penny, 2004). Much like the formal assimilation of protases and apodoses, this is not entirely surprising as it is not uncommon for pluperfects to come to be used to refer to hypothetical moments closer to the present. Bull (1960) and Rojo (1974), in Klein-Andreu (1991), noted the cross-linguistic tendency for pluperfects to "migrate forward" temporally as they become increasingly used in association with impossible situations temporally closer to the present as opposed to temporally more distant ones that actually did occur. The innovative use of *-ra* in protases without pluperfect past temporal reference marks the context where *-ra* truly began its variation as an Imperfect Subjunctive morpheme and it can be assumed that once it was interchangeable with *-se* in this context, it subsequently generalized to other imperfect subjunctive contexts. Klein-Andreu (1991) and Lapesa (1981) have argued that the addition of the periphrastic *haber + past participle* construction is what truly led to the blurring of the two forms, that is, of *-ra* and *-se*, as the survival of both ultimately relied on the strength of their imperfect subjunctive uses. Figure 2 summarizes the development of *-ra*.

**Figure 2: The Evolution of *-ra***




From a perspective of grammaticalization, the development of *-ra* and *-se* could be explained using the concept of the cline, a pathway composed of "links", much like a chain, that represent the various phases of change (Hopper & Closs Traugott, 2003). Since the two forms have unique origins, the question would become whether they are two forms that have truly merged to move along the same cline, albeit at different points, or if they are moving along their own parallel clines. Given that the two forms followed similar pathways to initially become Imperfect Subjunctive inflections (via the conditional sentence), it could be argued that they now represent two grams moving along the same cline. If this were to be the case, *-ra* would in theory represent a less-advanced grammaticalizing subjunctive form since it entered the chain long after *-se.* Additionally, if in this model the grammaticalization of both Imperfect Subjunctive morphemes behaved similarly, meaning they followed similar trajectories once established as Imperfect Subjunctive inflections, the generalization of *-ra* could mirror the generalization of *-se* as it assumed its imperfective, as opposed to pluperfect, reading. Testing this hypothesis would

require examining written texts for variation between *-se* and the Latin Imperfect Subjunctive *-re*. Such an analysis is outside the scope of this project; however, the present analysis may be able to shed light on the behavior of *-ra* as an Imperfect Subjunctive morpheme gaining frequency, which could provide a point of comparison for said analysis of *-se.*

Regardless of their potential degree of grammaticalization and, for now, putting aside any consequential nuances of meaning this may produce, in modern usage these morphemes are generally held to not show any major functional differences in subjunctive contexts (Penny, 2004). Butt & Benjamin (2011) suggest that they are completely interchangeable, except in contexts where *-ra* may appear with one of its historical values (pluperfect indicative or in the apodoses of conditional sentences). However, others (Lunn, 1989 and Luquet, 2004) have challenged the notion that, historically and in modern usage, the two forms carry equivalent semantic values in imperfect subjunctive contexts. A brief discussion of these non-subjunctive uses, although uncommon, and of the supposed semantic division between the two morphemes is merited as they have been frequently discussed in prior *-ra*/*-se* variation work.

## 2.2 The Debated Allomorphy of *-ra* and *-se*

Although *-ra* and *-se* are both classified as Imperfect Subjunctive morphemes and are most commonly used as such, they are not considered true allomorphs by all (Luquet, 2004; Lunn, 1989). As previously mentioned, even in modern usage, *-ra* has retained the ability to function as a variant of the Pluperfect Indicative. Prescriptively, *-se* does not share this ability, being "rejected by grammarians" in these contexts (Butt & Benjamin, 2011, p. 222). However, as attested by corpus data (Lunn, 1995; Sorenson, 2016), *-se* can appear in these contexts in actual language use, presumably by analogy of its variability with *-ra* in other contexts. Examples (6)

and (7), taken from Butt & Benjamin (2011) and Sorenson (2016), illustrate occurrences of Pluperfect Indicative *-ra* and *-se*. The translations were also taken from these sources.

(6) Fue el único rastro que dejó en el que **fuera** su hogar de casada por cinco horas.

´It was the only trace she left in what **had been** her marital home for five hours.'

(Butt & Benjamin, 2011, p. 229)

(7) Hace poco más de dos meses que la persona con la que **hablase** en Madrid descargaba una historia de recriminaciones contra la canciller alemana.

´It was just over two months ago that the person with whom I **was speaking** in Madrid unleashed a torrent of recriminations against the German chancellor.´

(Sorenson, 2016, p. 832)

As seen in the above examples, this specialized use of Imperfect Subjunctive forms is known to be restricted to subordinate clauses. Penny (2002) notes that this use is specific to relative clauses and Lunn (1989) has found that its use is also constrained to clauses which present information that is already known. Butt & Benjamin (2011) and Penny (2004) have both noted that its use is also particular to literary and journalistic prose. The restriction of indicative *-ra* (and only by analogy *-se*) to subordinate clauses exemplifies a retention of its prior use. It is common for grammaticalizing forms to not completely lose their former uses or distribution patterns as they gain new ones (Hopper & Closs Traugott, 2003; Torres Cacoullos, 2001). Klein-Andreu (1991) has argued that indicative *-ra* was highly constrained to subordinate clauses as the developing *haber* + *past participle* took over in main clause contexts, so it would follow that in modern usage this context would be the one retained.

As discussed above, even after the development of Conditional *-ría*, *-ra* retained its ability to convey conditional value in the apodoses of conditional sentences and came to be used with certain high-frequency modal verbs in non-apodoses contexts (i.e. *quisiera* = *querría* 'I

would like) (Penny, 2002). Prior to the 18th century, Conditional *-ra* was quite common and could be used with all verbs in conditional apodoses. With time, however, Conditional *-ría* supplanted *-ra* as the favored Conditional form in conditional sentences (just as *-ra* had done to *-se* in this context). In the present day, Conditional *-ra* continues to be used with only a few modal verbs, namely *quisiera, pudiera, debiera,* and *hubiera.* In Latin America, *pareciera* has also come to be used in this way in the formulaic literary construction *pareciera que* 'it would seem that'. It is important to note that the other aforementioned verbs are also often used in other formal formulaic constructions that tend to denote courtesy or politeness (Butt & Benjamin, 2011; Lunn, 1995). It is not impossible for *-se* to appear in these contexts by analogy, but Lunn (1995) notes that using *-se* in this way is considered "obsequious and pompous" (p. 438). Luquet (2004) takes a stronger position regarding this use of *-se,* suggesting that such a case would be strictly ungrammatical. The one exception to this disfavored use of *-se* could be auxiliary *hubiese.* In general language use, the analytical Pluperfect Subjunctive can appear as a main clause Pluperfect Conditional or in the apodosis of counterfactual conditional sentences with either *-ra* or *-se*, as shown in (8), (9), (10) and (11).

(8) Gusto de saber tu historia, y más te **hubiera escuchado**, mas el día apresurado su curso acaba.

'I'm pleased to know your story, and I **would have listened** to you more, but the course of the day is quickly ending.'

AC, 17th cen.

(9) Don Álvaro no era superior a su siglo, y en cualquiera otra ocasión, semejantes circunstancias no **hubiesen dejado** de hacer impresión profunda en su ánimo.

'Don Álvaro wasn't superior for his century, and in whatever other occasion, similar circumstances **would** not **have ceased** to leave a deep impression on his mood.'

EA, 19th cen.

(10) **Hubiérasme** así **obligado**...a no sospechar que mientes.

'In that way you **would have obliged** me, if I didn't suspect that you're lying.'

AC, 17th cen.

(11) Si aquella tarde su tío no se lo sacaba de encima, Cornelio lo **hubiese destrozado**.

'If that afternoon his uncle did not take it down from on top, Cornelio **would have destroyed** it.

DLC, 20th cen.

Lunn (1995) notes that it is precisely this use (i.e. with auxiliary *haber*) where *-se* has been most conserved in dialects that use *-ra* almost exclusively in other Imperfect Subjunctive contexts.

In true subjunctive contexts one would logically think that *-ra* and *-se* may be considered equals, but this has been much debated. While authors such as Penny (2004) and Butt & Benjamin (2011) have asserted that in subjunctive contexts the two forms are completely interchangeable. Luquet (2004) and Lunn (1989) have challenged this notion, drawing upon the highly nuanced nature of the subjunctive itself as evidence that such a division could exist.

Pinpointing a common meaning for verbs encoded in the subjunctive has proven to be a challenge for linguistic scholars since the subjunctive can appear in a variety of syntactic environments and carry a range of semantic meaning. As defined by Ridruejo (1999), the subjunctive can express the eventuality or futurity of a situation not yet realized at the moment of speech, communicate the uncertainty or the desire of the speaker or communicate hypothetical situations. Further complicating its overall interpretation is the fact that there are contexts in which the subjunctive may be used, but it is not obligatory (Butt & Benjamin, 2011). To this point, Lunn (1989; 1995) adds that choosing the subjunctive often reflects a speaker´s attitude toward the verbal situation since, in addition to being able to appear in numerous non-obligatory contexts, there are many instances in which a speaker could rephrase their utterance to

communicate the same idea without using the subjunctive at all (see Lunn, 1995, p. 443 for examples). Finally, Lunn (1989) also presents the anomaly that the subjunctive can be used "to mark both information that is untrue or contextually unrealized. . .and information that is demonstrably true¨ (p. 688). In light of this semantic complexity, the possibility of different meanings for *-ra* and *-se* would not be inconceivable.

To account for its myriad uses, Lunn (1989; 1995) has suggested that nonassertion may serve as the most appropriate unifying definition for the subjunctive. This accounts for the aforementioned anomaly that speakers may choose to not assert information because it is either untrue or because it is old information that need not be foregrounded. In essence, she summarizes use of the subjunctive to mark content that "add[s] nothing true. . .or nothing new to existing information" (Lunn, 1995, p. 430). Lunn (1989) further suggests that within the framework of assertability lies the possibility of a scalar conceptualization of the subjunctive. This is, the assertion/nonassertion dichotomy is not truly binary. She draws upon a limited number of authentic examples to promote the notion that the variation between *-ra/-se* serves as evidence of this gradient of meaning. She suggests that, when analyzed in their discourse context, *-se* emerges as the less-assertive, more polite of the two forms and is the form that suggests lower likelihood of occurrence. She also argues that the abundance of *-ra* forms in main clause subjunctive contexts as well as indicative uses of *-ra* can still be accounted for within this framework. Maintaining that *-se* is the less assertive of the two forms, it would logically follow that it would appear less frequently in main clauses, the "domain of assertion" (p. 697). Furthermore, she asserts that marking information with *-ra* can serve to background its importance. She presents an analysis of examples from historic and journalistic texts which revealed that Indicative *-ra* tends to appear in subordinate clauses that add information that is

either unimportant or that is already known (a finding that is congruent with Klein-Andreu's (1991) historical data). She notes a particularly high occurrence of Indicative *-ra* in newspaper sections, such as the gossip column, that would have a consistent readership base and feature recurring subjects or in statements that repeat information from a headline (Lunn, 1995). Examples (12) and (13) illustrate such uses of Indicative *-ra* taken by Lunn (1995) from the popular magazine *Hola* (different issues published in 1985) (p. 433). Semantically speaking, then, using the analytical pluperfect indicative form with *haber* or a preterit verb in these contexts would promote the saliency and assertability of the added information. The translations are also from Lunn (1995).

(12) La pareja, que **se hiciera** famosa por interpretar el papel de marido y mujer en "El pájaro espino", es en la vida real un matrimonio feliz.

'The couple, who **became** famous for their role as husband and wife in "The Thorn Birds", is happily married in real life.

(13) Y al final, besó la bandera roja y gualda, que hace treinta años **besara** su padre el rey y que un día **bordara** su tartarabuela la Reina doña María Cristina.

'And, at the end, he kissed the red and gold flag that his father **had kissed** thirty years ago and that his great-great grandmother Queen María Cristina **had** once **embroidered**.

In a different attempt to conceptualize the division between all Spanish verb forms, Luquet (2004) organizes them in two categories: the realized and the unrealized. He placed the subjunctive verb forms within the latter. Going a step further, he draws upon a limited number of examples from literary texts to illustrate that, at least prior to the conclusion of Spain's Golden Age in the 17th century, *-ra* and *-se* were semantically contrastive even though they were temporal equals. Luquet sustains that *-ra* was the form selected to signal situations that were unrealizable or contradictory to reality. Contrarily, he proposed that verbs conjugated with *-se*, while still unrealized, conveyed a higher degree of possibility of realization and expressed doubt,

desire or the subjective evaluation of a situation. For Luquet, this semantic distinction faded, but has not been completely lost, maintaining that *-ra* still carries a stronger connotation of unrealizability than *-se*. Although this may seem contrary to *-ra´*s original indicative meaning, this nuance of unrealizability may be remnant of *-ra´*s initial appearance in conditional sentences that communicated impossible hypothetical actions of the past. Luquet hypothesizes that, with time, this distinction will eventually cease to exist.

Bybee *et al.* (1994) offer an explanation using the theory of grammaticalization that can account for the aforementioned complexity of the Spanish Subjunctive. If we return to the idea that grammaticalizing forms can represent links on a chain, this allows for grams moving along the same cline, but at different points, to not behave similarly, thus eliminating the need to unify all uses with one definition. Furthermore, as grams gain new meaning and extend into new contexts, their previous functions and contexts of use are not necessarily lost. This could account for the wide range of contexts and nuances now belonging to the Spanish Subjunctive. A parallel to *-ra* and *-se* may also be made here. If *-ra* and *-se* are examined through this lens, it could be posited that their perceived inequalities are due to their different locations on the Imperfect Subjunctive grammaticalization chain. As previously presented, *-se* acquired its Imperfect Subjunctive function first and was much later followed by *-ra*. For this reason, *-se* could represent a more advanced stage of Imperfect Subjunctive development. Perhaps the residual contexts that seem to favor *-se* reflect domains that represent more advanced stages of Subjunctive grammaticalization that *-ra* has not yet fully developed.

## 2.3 Other Observations of *-ra*/*-se* Distribution

The semantics of the two forms is but one avenue down which *-ra*/*-se* research has gone. Sociolinguistic factors which are known to influence linguistic variation, such as age, gender,

social class and regional identity, have been analyzed but have demonstrated mixed results across the examined varieties (Williams, 1982; Serrano, 1996; Martínez, 2001; Branza & van Heuven, 2005; Kempas, 2011; Sorenson, 2016). Usage of -ra/-*se* in various genres of texts has also been studied, but, as with the sociolinguistic factors, mixed results emerge. Diachronically, Hanna (2012) found that literature was more innovative than other written texts (i.e. academic or scientific), but Williams (1982) found that, synchronically, regional variation or author stylistic choices may have a stronger influence. As previously mentioned, the majority of studies have revolved around the syntactic distribution of the two forms in dependent clause contexts. While there has also been non-uniformity of results across the examined varieties (both synchronically and diachronically), several studies have found that adverbial clauses have been more innovative than adjectival or nominal clauses (Wilson, 1983 in Martínez, 2001; Ramírez Luengo, 2001; Branza & van Heuven, 2005).

## 2.4 Conclusion

Documenting the relative frequency of the two forms in various regions at different points of time illuminates but one small component of this topic. The distributions of *-ra*/*-se* found in previous studies do suggest that syntactic context may very well have had a role in the gradual diffusion of *-ra* over time. However, questions still remain about *why* certain contexts initially attracted *-ra* to propel its generalization through various clause types. Since variation is multifaceted, observing additional (extra)linguistic factors through the statistical analysis of a larger corpus could provide more reliable data that goes beyond relative frequency alone to hypothesize a path of generalization of *-ra* over a given period of time. The present study, consequently, has been designed in an attempt to contribute to this discussion by addressing these needs.

# Chapter 3 - Analyzing *-ra*/*-se* Variation

## 3.0 Making Sense of Linguistic Variation

Although the average speaker of any language may not be aware that in day-to-day language use multiple linguistic forms are readily available within his or her linguistic repertoire to convey a particular message, such variation is an inevitable characteristic of human language (Tagliamonte, 2006). Simultaneously, as suggested by Labov (1986), selecting one form over another can facilitate a speaker's ability to linguistically index his or her identity as particular variants come to be used more frequently and systematically among social groups. Furthermore, as summarized by Tagliamonte (2006), in situations where overt identity construction may not seem influential in the selection process, it has been shown time and time again that linguistic variation is indeed quite patterned even though selection of competing forms may seem entirely random. These ideas together suggest that underlying internal and external factors influence, or constrain, a speaker's selection of forms. Out of the need to provide explanations for these observed phenomena, variationist methodological approaches have been developed over the last half century.

Statistical analyses have long been used to make sense of data patterns. However, given that language data "has heterogeneity with contextually conditioned 'order' to it as well as innumerable blank regions" (Tagliamonte, 2006, p. 130), traditional statistical modeling programs have not been able to fully account for nor faithfully explain the patterning of variation within naturalistic language samples. It was for this reason that variable-rule programs have been developed to take into account the uneven distribution of forms inherent in natural language use. The underlying logic of the program helps minimize the effects of such naturally occurring conditions to provide a more accurate analysis. Additionally, given the tendency for factor

groups to interact, variable-rule analysis programs have been designed to test various combinations of factor groups. In so doing, the program can select which scenario produces the most favorable and statistically significant conditions for selection of the dependent variable under study. It should be kept in mind, however, that variable-rule analysis still has its limitations and is only one tool that can be used. Nonetheless, it is invaluable to researchers who seek to ascertain a possible explanation for the patterning of language use in both synchronic and diachronic studies, such as the one at present.

## 3.1 Diachronic Variationist Analysis

As language change entails the shifting distribution of competing forms over time, a variationist approach can be applied to diachronic data to hypothesize the pathways of this change. By analyzing statistical data detailing the relative impact of an array of relevant (extra)linguistic independent factor groups from multiple centuries, a theory of a form's underlying grammar may be established for each century and then compared to determine the innovative contexts in each century and which factors most constrained the selection of the dependent variable over time. These contexts can then, in turn, be proposed as having been the most influential as the form generalized over time (Tagliamonte, 2006).

Previous work on *-ra*/*-se* variation has not been carried out in this way. With the exception of McKinnon's (2018) synchronic study of *-ra*/*-se* variation in Catalonian Spanish, variable-rule analysis has not been utilized. This study, therefore, addresses the need to examine *-ra*/*-se* variation with an approach that relies on the statistical weighting of combinations of factors instead of simply comparing relative frequencies in individual contexts (i.e. syntactic context alone). Expanding analysis to include other factor groups, which will be discussed in Chapter 4, may not only provide a clearer picture of the expansion of *-ra* at the expense of *-se*,

but it may also contribute to the debate about the (lack of) equivalence of the two morphemes. Strong favoring of one form over the other in certain contexts could be indicative of nuances the two forms may not have historically shared.

Before proceeding any further, it should be noted that the *-ra*/*-se* variation does indeed lend itself to variationist analysis. As defined by Tagliamonte (2006), corpus data compiled to examine the use of particular variants through variable-rule analysis must meet three criteria, which the data in this study fulfill. First, there must be a clear choice of analogous variants. By the 17th century (the first century included in this study), *-ra* and *-se* were both known to be Imperfect Subjunctive morphemes, and *-ra* could be used in all contexts where *-se* had previously occurred (Lapesa, 1981; Penny, 2002; Pharies, 2015); hence, they may be analyzed as competing linguistic variables. Second, the selection of the variants must seem unpredictable and "haphazard based on known parameters" (Tagliamonte, 2006, p. 131). Variation between *-ra*/*-se* is present in all works included in this study, it occurs within the same sentence at times and also appears in any context that requires the Imperfect Subjunctive. Third, the selection must be recurrent in language use. While the Imperfect Subjunctive itself is not very frequent, there are still enough obligatory contexts in which it is utilized that a sizeable corpus can be constructed. Furthermore, in historic data, as opposed to modern, synchronic analyses (see Sorenson [2016] for example), use of either variant is not so infrequent as to yield a token count that is too limited for analysis. The next sections outline the process followed to compile the corpus used for this study.

## 3.2 Corpus Construction

### 3.2.1 Selection of Corpus Texts

As most succinctly presented by Sorenson (2016), *-ra*/*-se* variation is not unique to any one variety of Spanish. That is, all dialects have both morphemes at their disposal, albeit with varying frequencies of occurrence. Furthermore, as shown by the search of the *Corpus de español* (Davies, 2002) presented in Table 1, *-se* was seemingly still the preferred morpheme in the 16th century when Spanish colonization of the Americas began (at 70% of the tokens). As such, it would be possible to suggest that whichever mechanisms were already at work in Peninsular Spanish when it spread throughout the Americas continued to propel *-ra* forward as the dominant Imperfect Subjunctive morpheme. Although a more complete analysis of historical regional *-ra*/*-se* variation is needed, Peninsular Spanish has been selected for analysis as it is the source language of Latin American varieties of Spanish and due to the quantity of easily accessible documents from the last several centuries.

Literary texts have been selected for analysis as they proved to be innovative (i.e. more favoring of *-ra*) in previous work on *-ra*/*-se* variation (Hanna, 2012). Ten Peninsular literary texts (a combination of dramatic texts and prose) were used to compile a corpus of *-ra*/*-se* tokens. As the aim of this study is to identify how *-ra* usage expanded over time, separate corpora were constructed for the 17th and 19th centuries. Analyzing texts from alternating centuries is appropriate as mechanisms of change tend to work slowly on linguistic systems (Bybee *et al.*, 1994). Analyzing non-sequential centuries, in theory, permits a sharper contrast to emerge between the characteristics of usage over time, therefore better clarifying the contexts of change (Copple, 2009). Additionally, the 17th century is an ideal point of departure as it was by

this century that *-ra* is believed to have assumed its Imperfect Subjunctive meaning (Lapesa, 1981).

Three dramas and two novels were selected for each century. While all acts of each drama were included for token extraction, only portions of the novels were used so as to ensure similar token counts from each work. When possible, subdivisions of the book were utilized (e.g. books 1 and II of *El buscón*). In other cases, such as with *El Señor de Bembibre,* tokens were extracted from the first few chapters (enough that the overall count would be similar to the dramas). To minimize any generational differences that may have influenced the authors' choice of verb form, all of the texts selected for each century were written within 50 years of each other. Uncontrolled for, however, was the author's town of origin. Many were from central Spain, living most of their life in Madrid, but other regions of the country are also represented (see Tables 4 and 5). This may present a possible limitation as previous work has revealed that there are slight regional differences of *-ra*/*-se* usage within Spain (Kempas, 2011; Williams, 1982; Pollán, 2001; McKinnon, 2018). Since semantic aspects based on the texts' events as well as sociolinguistic factors were included for initial coding, understanding the plot was necessary. Thus, brief summaries of each text may be found in Appendix B.

#### 3.2.1.1 17th Century Texts

The texts selected for the 17th century belong to Spain's Golden Age and were written between 1608 and 1628. While the authors represent some of the most well-known writers of the time, the works selected are not necessarily considered canonical today. The selected dramas follow the standard conventions of the time, all having been composed in verse and featuring themes of love, social power and the code of honor (Parker, 1998). The selections of prose include the popular *novela picaresca* (Quevedo) and a lengthy short story (from Cervantes'

*Novelas ejemplares*). Both selections feature dialogue, but are otherwise told with first- and third-person narration, respectively. Unlike the three dramas, these texts do not include royalty. Table 4 summarizes the texts selected for the 17th century.

| Overview of 17th Century Texts | | | |
|---|---|---|---|
| **Genre** | **Work (with example citation code)** | **Author (with birthplace)** | **Word Count[7]** |
| Dramatic Texts | *Amor, honor, poder* (AHP), 1623 | Pedro Calderón de la Barca (Madrid) | 15,390 |
| | *El amor constante* (AC), 1608 | Guillén de Castro (Valencia) | 19,100 |
| | *La vengadora de las mujeres* (VDM), 1613-1620 | Félix Lope de Vega (Madrid) | 15,000 |
| Novels | "La gitanilla" from *Novelas ejemplares* (GIT), 1610 | Miguel de Cervantes (Alcalá de Henares) | 21,204 |
| | *El buscón* (BUSC), 1628 | Francisco de Quevedo (Madrid) | 23,736 |

**Table 4: 17th Century Texts**

#### 3.2.1.2 19th Century Texts

Texts selected from the 19th century were written between 1806 and 1846, a turbulent time for Spain due to the French invasion, occupation, and subsequent political unrest. Literary tendencies of the first half of the 19th century are considered to align with the transition from Neo-Classicism to Romanticism (King, 1962). Romanticist tendencies may be seen in the

[7] The word count for each dramatic text were based on the word count generated by Microsoft Word. Word counts for each novel were calculated by averaging the number of words for three pages of text and then multiplying that number by the number of pages utilized.

adventurous plots of certain works (*Don Álvaro* and *El señor de Bembibre* in particular). The linguistic characteristics of this literary style will be incorporated into the discussion of the results in Chapter 5. The dramatic texts of this period are linguistically more straightforward than those from the 17th century, featuring shorter exchanges of dialogue and fewer lengthy monologues. They were also written in verse, prose or a combination of the two styles. Characters featured were no longer royals but rather members of the emerging bourgeoisie. However, themes still centered on love and power. The novels selected for this century differ in that one simply tells a fictitious story (Gil y Carrasco) while the other (Calderón) captures *costumbrista*, or ordinary, interactions between Andalusian bar regulars and expositions on local "heroes". In parallel to the 17th century, these novels both include dialogue and feature third-person and first-person narration, respectively. Table 5 presents a summary of the texts utilized from the 19th century.

| **Overview of 19th Century Texts** | | | |
|---|---|---|---|
| **Genre** | **Work (with example citation code)** | **Author (with birthplace)** | **Word Count** |
| Dramatic Texts | *Don Álvaro o la fuera del sino* (DA) 1835 | Duque de Rivas (Córdoba) | 23,000 |
| | *El afán de figurar* (ADF), 1831 | José María de Carnerero (Madrid) | 17,050 |
| | *El sí de las niñas* (SDN), 1806 | Leandro Fernández de Moratín (Madrid) | 18,500 |
| Novels | *El Señor de Bembibre* (1843) SDB | Enrique Gil y Carrasco (Valencia) | 23,842 |
| | *Escenas Andaluzas* (EA), 1846 | Serafin Estébanez Calderón (Málaga) | 22,661 |

**Table 5: 19th Century Texts**

### 3.2.2 Extraction

All occurrences of Imperfect Subjunctive forms (including *haber* used to form the Pluperfect Subjunctive) were initially extracted for analysis. This includes the supposed "conditional" uses of *-ra*, discussed above. Although it is common in variationist analysis to only extract variants with the same function (in this case, truly imperfect subjunctive), broadening the variable envelope may actually provide a better image of how a form's use spread to novel contexts (Copple, 2011). The "conditional" uses occur in environments where *-se* had also historically been used (i.e. conditional apodoses) and are therefore considered variable contexts. Table 6 shows the total count of *-ra*/*-se* tokens extracted from each work in both centuries.

| **Token Counts from Corpus Texts** | | | | |
|---|---|---|---|---|
| | 17th Century | | 19th Century | |
| Dramatic Texts | *Amor, honor, poder* | 170 | *Don Álvaro* | 71 |
| | *El amor constante* | 119 | *El afán de figurar* | 64 |
| | *La vengadora de las mujeres* | 108 | *El sí de las niñas* | 87 |
| Novels | *El buscón* | 179 | *El señor de Bembibre* | 91 |
| | *La gitanilla* | 173 | *Escenas andaluzas* | 124 |
| **Total** | | **751** | | **437** |

**Table 6: Total *-ra*/*-se* Token Counts by Century and Text**

Once collected, these tokens were coded for a number of (extra)linguistic factors which will be discussed in the following chapter.

# Chapter 4 - Coding of Factor Groups

## 4.0 Introduction

Once all tokens had been extracted, with their surrounding context, they were coded for a number of (extra)linguistic factors. Given that prior research has primarily focused on syntactic context, this study aims to include other linguistic dimensions. Factors were selected for coding in order to investigate text-specific dimensions, linguistic aspects related to the token itself as well as aspects particular to the context in which the token occurred. Lastly, as the Imperfect Subjunctive can refer to actions with different points of temporal reference as well as actions with varying degrees of realization, related factor groups were also included for coding.

## 4.1 Textual Factors

Tokens were coded for the genre in which they appeared (novel or drama) as well as for the author of the text from which they were taken. Since both dramas and novels incorporate dialogue, this was also added as a factor group (coded binarily as 'dialogue' or 'narration'). Furthermore, Serrano (1996) and Williams (1983) have found that age, level of education, and gender may influence *-ra*/*-se* variation. For that reason, sociolinguistic factors, such as gender, and social class, were also considered at the onset of the coding process. Gender of the speaker/character in tokens taken from dialogue was coded. Inclusion of social class would have been interesting; but given that these texts represent various communities (royal courts and small villages), both in and out of Spain, this category was excluded as it was too difficult to confidently assign each character to a social grouping. It is also worth noting that this measure would only offer a possible indication of social class effects since any choice in form made to reflect actual speech of the time period would have been subjectively made by the text's author. Although the same may be said for gender, at least this category could be confidently assigned.

## 4.2 Linguistic Factors

The linguistic factors included for coding may be divided into three groups: those which are related to the verb encoded in *-ra*/*-se*, those which are related to the syntactic context in which the token occurred and those which pertain to uses of the Imperfect Subjunctive as a manifestation of tense and mood.

### 4.2.1 Verb-related Factors

Originally included for coding were the grammatical subject of the verb (1st, 2nd, 3rd person) and its polarity (affirmative or negative as determined by the presence of '*no*' or other negative word before the verb to negate the action). Additionally, frequency of form has been shown to affect language change (Bybee, 2006). For this reason, counts of all verb types (represented by each infinitive) were generated to classify all verbs binarily as "high-frequency" or "low-frequency". Verbs which occurred a total of 15 times or more were considered highly frequent[8].

### 4.2.2 Clause-related Factors

The use of the Imperfect or Pluperfect Subjunctive in Spanish is often influenced by the semantic and syntactic relationships established by the overall meaning of the clauses of the complex sentence in which they occur (although some main clauses also utilize these verb forms). Since subordinate clauses, which may be divided into three types (described below), are a prototypical context for the subjunctive, observing *-ra*'s degree of generalization in each clause type is of interest, especially since dependent clauses show varying levels of subordination

[8]The decision to use 15 tokens as the starting point for 'high frequency' was arbitrary. However, in the 17th century corpus, which was used to make this decision, verbs with 15 tokens or more generally patterned together (except for *dar* 'to give'). The list of verbs coded as high-frequency may be found in Chapter 5.

(Hopper & Closs-Traugott, 2003). For this reason, the syntactic function of the clause in which each token occurred was included in the coding schema[9].

The Imperfect and Pluperfect Subjunctive may occur in both independent and dependent clauses. At the onset of coding, independent clauses were divided between the apodoses of conditional sentences, underlined in (14), and other main clauses, as in (15).

(14) Y si no fuera por vuestra espada y favor, <u>mi carrera sin honor ya **estuviera** terminada</u>.

'And if it weren't for your sword and favor, <u>my career **would** already **be** over without honor</u>.'

DA, 19th cen.

(15) ¡Oh quién **fuera** el dichoso que la vida te diera!

'Oh if only I **were** the fortunate one who could spare your life!

AHP, 17th cen.

The remaining dependent clauses were divided among four categories: nominal clauses, adjectival clauses, adverbial clauses and the protasis of conditional sentences. In each of these clause types, the content of the clause serves a different syntactic function. Hualde *et al.* (2010) summarize the clause types as follows. Nominal clauses function as nouns, serving as the direct

[9] The polarity of the clause was also included for coding, but was ultimately excluded from analysis due to the extreme low frequency of negative clause heads (N = 27/521 or 5.2%). Although it had also been coded for, the semantics of the clause was also excluded due to the non-orthogonality of this category. As Tagliamonte (2006) emphasizes, this is problematic since factor groups must not represent a subcategorization of another. The semantic class of a clause inherently correlates with its syntactic type (e.g. adverbials may be divided into purposive, temporal, etc.). A possible solution to this problem would have been to only include the semantic groups in the multivariate analysis, but when this was attempted, interference of an unknown origin skewed the results.

object of a verb (16), as the subject of an impersonal statement (17) or as the noun phrase complement of a prepositional phrase (18).

(16) Mandaron los doctores que, por nueve días, no **hablase** nadie.

'The doctors ordered that, for nine days, no one **should talk**.

BUSC, 17th cen.

(17) ¿Cómo es posible que yo la **abandonase**... ¡criatura!..., en la situación dolorosa en que la veo?

'How is it possible that I **abandoned** you…child!..., in the painful situation in which I see you?

SDN, 19th cen.

(18) …insistió en que la **acompañase** el cobrador de las rentas del convento.

'[he] insisted [on] that the rent collector of the convent **accompany** her .'

DA, 19th cen.

Adjectival clauses describe an entity present in discourse. They provide information that can help distinguish one entity from another, or they can simply give additional information to further describe a particular entity which is already known. The subjunctive is used in these clauses to modify nouns that may possibly exist (19) or nouns that are known to be non-existent (20).

(19) Pajaritas del aire **que apetecieras** las tendrías

'Little birds in the air that **you may fancy** you would have'

SDN, 19th cen.

(20) No había diablos que me **moviesen** de un lado.

'There weren't demons that **could move** me from one side.

BUSC. 17th cen.

Adverbial clauses share a relationship with the verb of the main clause. When these relationships relate two actions that are dependent but could not occur simultaneously, the subjunctive is often utilized in Spanish. In (21), for example, due to the uncertainty of when Simón would notify Don Carlos of when it would be safe for him to move from his hiding spot, this action is encoded in the Imperfect Subjunctive.

(21) Le encargué que no subiera hasta que le **avisara** yo, por si acaso había gente aquí, y usted no quería que le viesen.

'I charged him to not go up until I **notified** him, in case there were people here, and you didn't want them to see him.'

SDN, 19th cen.

Even though the protasis of a conditional sentence is a type of adverbial clause, these clauses were coded separately due to their historical significance in the subjunctification of *-ra* (Lapesa, 1981; Klein-Andreu, 1991). Conditional sentences are composed of a formula which proposes a condition established in the protasis (the *si* 'if' clause) that, if realized, would facilitate the result expressed in the apodosis, as seen in (22). Here, the Infanta is implying that Enrico is unaware of the heaviness of his request and suggests that if he were, he would not be making it due to the weight it would put on his conscience.

(22) Enrico, si tú **supieras** lo que a pedirme te atreves, sospecho, que te pesara.

'Enrico, if you **knew** what you're daring to ask of me, I suspect that it would weigh on you.'

AHP, 17th cen.

As stated above, dependent clauses, while all subordinate to a main clause, do not share equal degrees of subordination; this is, they can be understood as existing on a continuum of embeddedness. Ideas presented in discourse are not isolated; they depend on each other and their context to create meaningful exchanges. However, the ways in which these ideas may be

connected can range from simple juxtaposition (23) to full subordination (24) (López García, 1999). As can be seen in these examples, the same idea is communicated, an invitation to come to the party, but with different syntactic structures.

(23) I'm throwing a party on Saturday. You should come.

(24) I'd like you to come to my party on Saturday.

It has been cross-linguistically observed that languages develop in such a way that ideas tend to become increasingly reported with more closely connected clauses (López García, 1999; Hopper & Closs Traugott, 2003). This is, the tendency to use simple juxtaposition or coordination of independent clauses, as in (23), is gradually replaced by more complex sentences with dependent clauses, like that of (24). However, as stated above, not all dependent clauses were created equally; some are more dependent than others. Along this continuum of dependency, the relationships of syntactic clauses can be categorized as paratactic (independent) as seen in (25), hypotactic (dependent but not fully embedded) as in (26), or subordinate (fully embedded) as seen in (27) (Hopper & Closs Traugott, 2003).

(25) Si hubiera vivido diez y ocho siglos después o en nuestros días, la notara, fijara y ampliara por todas aquellas grandes provincias.

'If he had lived eighteen centuries later or in our days, he would take note of it, focus on it and expand it throughout all those great provinces.'

EA, 19th cen.

(26) Doña Beatriz le hizo un ademán lleno de nobleza para que se sentase.

'Doña Beatriz made a noble gesture toward him so (that) he would sit down.'

SDB, 19th cen.

(27) Me mandó <u>que saliera inmediatamente de la ciudad</u>, y fue preciso obedecerle.

'He ordered me <u>to immediately leave the city</u>, and it was necessary to obey him.'

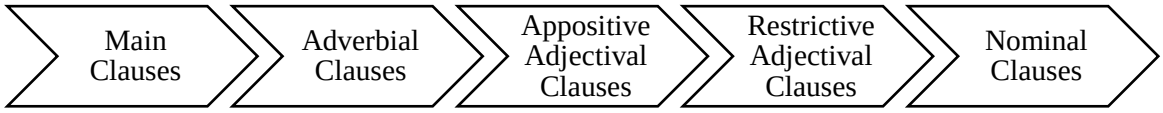
ADF, 19th cen.

In (25), commas and the coordinating conjunction *y* 'and' are being used to link three syntactically independent ideas together. A degree of semantic dependence can be noted, however, in the fact that the three underlined verbs cannot change order and still make sense. In the most independent of paratactic sentences, not only are the constituents syntactically independent, but they are also semantically independent in the sense that the ideas can be reordered and remain logical. In (26), a higher degree of dependency is present between the two clauses since *para que se sentase* cannot stand alone. However, it is not syntactically embedded, it can be moved to the beginning of the sentence (cf. *Para que se sentase, Doña Beatriz le hizo un ademán lleno de nobleza*), nor is it semantically necessary since it could be removed and leave a complete idea, albeit less specific. This contrasts with (27). Here, we see that the embedded clause, which is truly subordinate, is functioning as an argument within the noun phrase complement of *mandar*, a ditransitive verb that requires two complements. Unlike in (26), the subordinate clause cannot be removed (cf. **Me mandó y fue preciso obedecerle.*).

To more specifically apply these levels of dependence to the syntactic clause types included in this analysis, main clauses share paratactic relationships and are often conjoined by coordinating conjunctions (e.g. *and*, *or*, *but*). Adverbial clauses and appositive adjectival clauses, which only add information related to the referent, are considered hypotactic since they are more syntactically separate from their neighboring constituents but still depend on them since they cannot stand alone. Restrictive adjectival clauses, which provide additional information that serves to distinguish one entity from another, and nominal clauses are considered truly

subordinate due to being syntactically embedded within another syntactic constituent. However, restrictive adjectival clauses demonstrate less dependency in that they still only provide additional information, whereas nominal clauses are syntactically embedded and semantically necessary (Hopper & Closs Traugott, 2003). This implies that a scalar progression of clause types can be established from the most independent to the most dependent syntactic context, visually represented in Figure 2.

**Figure 3: Progression of Clause Dependence**




## 4.3 Functions of the Imperfect Subjunctive

As previously discussed, the Imperfect Subjunctive (and by extension the Pluperfect Subjunctive) does not impart one common meaning (Ridruejo, 1999). For this reason, temporal reference and degree of realization were included as separate factor groups in the coding process.

### 4.3.1 Temporal Reference.

Although the Imperfect Subjunctive is often thought of as the "past subjunctive", it ought to be noted that it does not always refer to a situation that occurred in the past. While this use does fall within its domain, the Imperfect Subjunctive is also used in contexts that refer to hypothetical situations that transcend the boundaries of time. This is, the Imperfect Subjunctive is used to hypothesize about events or circumstances that are ongoing. They could have begun in the past, but they still can be applied to the present or even refer to the foreseeable future. This can be seen in (28). In this aside, the King is lamenting the fact that his love interest, Nísida, does not feel for him what he feels for her and is plotting about how he could possibly woo her if

he only got the chance to see her again in private. The present is signaled by the use of the adverb *ahora* ´now´.

(28) Ah cómo ahora le **hablara**. si a solas hablar **pudiera,** que quizá la **enterneciera** si mis males le **contara**.

'Ah how I **would talk** to her now, if I **could talk** [to her] alone, perhaps I **could move** her if **I told** her of my misfortunes.'

AC, 17th cen.

In contrast, a truly past tense use of the Imperfect Subjunctive refers to an action that was bound to the past, as in (29). Here, the Condesa is reporting to the Baron what had occurred earlier that day when she had gone to see the government officials about a particular matter.

(29) Y le mandó que un informe le **presentase** al momento sobre el libro y el autor.

'And he demanded of him that **he present** a report on the spot about the book and the author.'

ADF, 19th cen.

### 4.3.2 Degree of Realization According to Context

Events encoded in the Imperfect Subjunctive can refer to situations with varying degrees of completion; this is, they may refer to actions that were realized, actions that could potentially occur or actions that are contrary to fact. These three categories were utilized for coding this dimension of the Imperfect Subjunctive.

In sentences such as (30), the verb encoded in the Imperfect Subjunctive refers to an action that was known to have occurred when it was reported but at a moment posterior to that of the verb in the main clause. When Lisardo reports to the other characters what Alejandro had asked him to do, it was already known to have happened; he did indeed take the hair ribbons as Alejandro had requested. However, since the taking of the ribbons occurred after Alejandro made the request, the Imperfect Subjunctive is still used.

(30) Alejandro me pidió que unas cintas **tomase** para hechizarte con ellas.

'Alejandro asked me to **take** a few ribbons in order to put you under a spell with them.'

VDM, 17th cen.

This use differs from a situation in which the action expressed in the Imperfect Subjunctive is only a possibility at the moment it is used. This may occur in past contexts in which a future action is expressed (31) or in hypothetical situations that could eventually be realized (32). In both cases, these events are speculation; it is not known if they will actually occur but are indeed possible.

(31) Preguntóle su madre que le **dijese** la verdad**.**

Her mother asked her to tell the truth.

GIT, 17th cen.

(32) Quisiera sólo que **se explicase** libremente acerca de nuestra proyectada unión…

'I'd only like for her to freely explain herself [her feelings] about our future union…

SDN, 19th cen.

In counterfactual statements, the situation presented in the Imperfect Subjunctive contradicts reality by either proposing an alternative reality, as in (33), or by negating what was actually known to be true, as in (34).

(33) ¡Si **vivieran** mis tres difuntos!...

'If my three deceased husbands **were alive**!

SDN, 19th cen.

(34) Si **no os hubiera hallado** en el camino, las nuevas me volvieran a Ferrara.

If **I had not found you** on the way, the news would send me back to Ferrara.

VDM, 17th cen.

In both cases, the action expressed in the Imperfect Subjunctive was not the reality and could never be so. The poor widow's husbands will not come back to life and Alejandro did run into his servants on his way to visit the Infanta. Decisions regarding these two factor groups were made through a close reading of each text to determine what the character or narrator knew to be true at the moment the Imperfect Subjunctive was used.

## 4.4 Selection of Factors for Analysis

Once all factor groups had been coded, an initial analysis of the distribution of *-ra* and *-se* was conducted to make decisions about recoding and to select which factor groups would be included in the multivariate analysis. It is imperative that factors not overlap (Tagliamonte, 2006); observing distributions through cross-tabulation is one measure that can help identify potentially problematic combination of factors. As will be explained below, several factor groups were not selected for inclusion in the multivariate analysis for various reasons. The 17th century data was used as the basis for these decisions.

### 4.4.1 Excluded Factor Groups

Upon closer inspection, several factor groups were deemed incompatible for inclusion in the multivariate analysis. Of the textual factors, author and genre were found to overlap; this is, the three dramatists and the two novelists from the 17th century patterned similarly, reflecting the same division observed in the genre category. Author was therefore excluded. Genre was also excluded as it overlapped with the dialogue factor group. Since all tokens taken from the dramas occurred in the actors' lines, they were all coded as dialogue. The dramas, therefore, did not contribute any non-dialogue tokens. Similarly, gender of the speaker was also excluded since it pertained only to the tokens taken from dialogue. Of the verb-related factors, subject (divided between 1st, 2nd and 3rd person) was excluded due to overlap with dialogue. The second person

tokens only appeared in dialogue whereas the first and third could appear in either dialogue or narration (recall that one novel in each century had a first-person narrator). Lastly, neither of the Imperfect-Subjunctive-related factors were included in the multivariate analysis since they were also found to interfere. Non-past temporal reference was highly infrequent in narration (N = 7/521, or 1.3%). The degree of realization was also excluded as it was found to interfere with several other groups, such as temporal reference and syntactic clause type (e.g. realized cannot be non-past and adjectival clauses are rarely counterfactual). The remaining factor groups included in the multivariate analysis for each century were: verb polarity, verb frequency, syntactic category and dialogue/narration. Chapter 5 presents the results of these analyses.

# Chapter 5 - Constraints on *-ra*/*-se* in the 17th and 19th Centuries

## 5.0 Introduction

This chapter presents the results of the multivariate analyses of the use of *-ra* and *-se* in the 17th and 19th centuries. The initial findings related to the overall distributions of the two forms, especially their distribution across main and dependent clauses, will be discussed first. Then, a detailed discussion of the factor groups which were selected as significant by the variable-rule analysis will follow. Each century's data will be discussed separately; however, the 17th century will be used as a reference point to track the changing constraints on the use of *-ra* in the 19th century. Finally, from these analyses, a broad conclusion regarding the nature of this linguistic change will be presented.

## 5.1 Relative Frequency of *-ra* and *-se*

At the onset of analysis, all occurrences of *-ra* and *-se* in all syntactic contexts were included. Table 7 shows the relative frequency of the two forms in each century.

| **Relative Frequency of *-ra*/*-se* by Century** | | | | | | |
|---|---|---|---|---|---|---|
| | ***-ra*** | | ***-se*** | | **Total** | |
| | N | **%** | N | **%** | N | **%** |
| **17th Century** | 462 | **62** | 288 | **38** | 751 | **100** |
| **19th Century** | 281 | **64** | 156 | **36** | 437 | **100** |

**Table 7: Relative Frequency of *-ra*/*-se* in the 17th and 19th Centuries**

On the surface, it would appear that use of the two forms hardly changed. However, as the following analysis will reveal, this was not at all the case. While the relative frequency did not change much, there were significant changes in the contexts of use for *-ra*, underscoring again

the importance of submitting such variation to more detailed analysis that goes beyond relative frequency to fully understand how it develops over time.

## 5.2 Independent Clauses and the Generalization of *-ra*

Upon closer inspection of the distribution of *-ra* and *-se* across clause types, it became apparent that verbs employed in the apodoses of conditional sentences and in other main clause contexts almost categorically appeared with *-ra,* at 99% of the tokens in both centuries. Interestingly, as revealed by the overall token counts, the total use of both forms decreased from the 17$^{th}$ to the 19$^{th}$ century (N = 751 and 437, respectively); however, the proportion between the total use of both forms in independent and dependent clauses remained strikingly similar, at roughly 30% to 70%, respectively, in each century. This signifies that the decrease itself was not linked to *-ra* or *-se*'s use in either of these clause types.

The fact that subjunctive *-ra* appears with a much higher frequency than *-se* in independent clauses should not come as a surprise when the nature of grammaticalizing forms is considered. Recall that, in accordance with many cross-linguistic observations, pre-existing grams that acquire new functions tend to do so in main clauses (Bybee *et al.,* 1994). As we previously saw, conditional apodoses served as the entry point for non-indicative *-ra* and came to be its favored context of occurrence for many centuries. With an established irrealis value gained in this context, non-indicative *-ra* could have then begun to spread to other independent clauses that shared a hypothetical reading, such as those with *ojalá* used to express laments about what could have occurred in the past. As such, it logically follows that these are the most favorable syntactic contexts for subjunctive *-ra* in the data set, further confirming Bybee *et al.*'s (1994) observation that innovative forms first generalize in main-clause contexts before permeating subordinate ones.

Such strong favoring of *-ra* in main clauses would interfere with the variable-rule results by giving artificial representations of the impact of the factors where these tokens were included. For this reason, these contexts were deemed non-variable and, consequently, excluded from the final multivariate analysis. This is not to say, however, that this finding is in no way unimportant. On the contrary, it allows us to propose that non-indicative *-ra* generalized in independent clauses (before the 17th century) prior to spreading into dependent clause contexts. Of interest for the present analysis, then, is the spreading of *-ra* through dependent clause contexts as this indicates more advanced stages of grammaticalization as usage of *-ra* became more prevalent in subordinate contexts as a subjunctive gram. Let us now turn to the results of the subordinate clause analysis.

## 5.3 Dependent Clauses and the Generalization of *-ra*

Before examining the results of the multivariate analysis, let us return to a brief discussion of how *-ra* was able to move beyond highly formulaic conditional sentences to be used in other subordinate subjunctive environments, where it has now all but replaced *-se* in modern usage. As mentioned in section 2.1.2, *-ra*'s move from the apodosis to the protasis allowed it to begin to vary more freely with *-se*, but its presence alone in the protasis was most likely not the only factor that caused its eventual spread to other subordinate subjunctive contexts. Klein-Andreu (1991) appeals to the fact that we have not yet seen a similar spread of the Conditional beyond the protasis in dialects where this is a possibility. Rather, she suggests that *-ra*'s ultimate success as a subjunctive form available in various contexts (beyond the conditional sentence, especially the apodosis) was due to Indicative Pluperfect *-ra*'s confinement to particular subordinate contexts. This, in conjunction with its remote temporal reference, made it a good candidate for reanalysis. Unlike other verb tenses (simple past, present, etc.), the

pluperfect is a tense which seemingly occurs more frequently in subordinate clauses; indeed, in Klein-Andreu's corpus of 14th century Old Spanish, only 3/184 Pluperfect Indicative tokens, or 1.6%, were in main clauses (p.172) [10]. As such, the pluperfect could be taken as a generally less-assertive tense since, in order to establish the "past of the past" temporal relationship, it depends upon another action, which would assumingly be the more asserted of the two since it is temporally closer to the present. Klein-Andreu (1991) also found that Pluperfect Indicative *-ra* tended to occur more frequently in pragmatically backgrounded subordinate clauses while *haber + past participle*, the innovative form, was more frequent in foregrounded subordinate clauses[11], further distancing *-ra* from its original assertive indicative value (p. 173). Thus, according to Klein-Andreu (1991), *-ra*'s growing association with low assertion in subordinate clauses, the prototypical domain of the subjunctive (Bybee *et al.*, 1994), is what allowed it to gain subjunctive status in subordinate clauses and move beyond the conditional sentence, edging out *-se* along the way. In essence, *-ra* began to do to *-se* what *haber + past participle* had done to *-ra*. Evidently, this was still a change in progress in both of the analyzed centuries.

After removing the independent clause occurrences of *-ra* and *-se*, 521 tokens from the 17th century and 307 from the 19th century remained for further analysis. As seen below in Table 8, dependent clauses were still variable in both centuries. While this context, when viewed alone, shows greater distributional differences in usage of *-ra* and *-se*, the change in relative frequency between the two centuries is not statistically significant (*p-value*: 0.24). However, the fact that

[10] It would be interesting, however, to compare the distribution of the Pluperfect Indicative across clause types in other texts to verify this finding.

[11] Klein-Andreu (1991) coded pragmatically backgrounded contexts as those which are negated, refer to a state of being, existing, or a possession, or describe another entity (i.e. relative clauses). Foregrounded contexts involved animate referents or reference to main events in the story.

there is an increase in *-ra* points to its changing status as a morpheme in competition with *-se* in subordinate contexts. Once more, this highlights the need to subject such variation to multivariate analysis to find the specific contexts where change occurs.

| **Distribution of *-ra*/*-se* in Dependent Clauses in the 17th and 19th Centuries** | | | | |
|---|---|---|---|---|
| | **17th Century** | | **19th Century** | |
| | **N** | **%** | **N** | **%** |
| ***-ra*** | 236 | **45** | 152 | **49.5** |
| ***-se*** | 285 | **55** | 155 | **50.5** |
| | 521 | **100** | 307 | **100** |

**Table 8: Relative Frequency of *-ra*/*-se* in Dependent Clauses**

The slight overall increase in the usage of *-ra*, only 4.2%, between the two centuries is somewhat surprising. These findings are similar to the distributions taken from the *Corpus de español* presented in Table 1, which show an increase of only 7% between the two centuries (Davies, 2002). Linguistic change is slow, but given the stark contrast between the two forms in 20th century usage and beyond (see Tables 2 and 3 in section 1.0), a greater shift between the two centuries might be expected. However, when thought of in terms of the corpus source texts utilized (novels and dramas), several considerations may be offered to explain the relatively small increase in *-ra* seen in the 19th century.

First, Guzmán Naranjo (2017), citing Gili Gaya (1983), has suggested that competing forms may develop associations with particular modes of language use (i.e. speech versus writing). In this case, if *-ra* were much more prevalent in spoken language, it could follow that *-se* became more favored in written contexts. Secondly, Neo-Classicism and Romanticist tendencies were in vogue in Spain during the first half of the 19th century when the texts included

in the corpus were written. Spanish Romanticists, in particular, were known to reach back to the Middle Ages for inspiration, both in content and in features of language use (Pharies, 2015). For this reason, it would be logical that *-se* may come to be used with more frequency in writing for stylistic purposes to evoke older patterns of usage. Even so, we do not see a total reversion toward *-se*, but rather an image of language use that cannot be easily extrapolated to describe naturally occurring patterns of usage of the time in other social contexts. Even with those literary tendencies present, *-ra* did still gain some ground in the 19th century. As such, the question becomes which contexts were most innovative for *-ra* in the 17th century, and how might the factors constraining its use have changed in the 19th century? Let us now turn to the results of the multivariate analysis to begin to delve into the answers to these questions.

## 5.4 17th Century Multivariate Analysis

Recall that of those factors which were originally included in the coding schema, only four were included in the variable-rule analysis carried out in Goldvarb X (Sankoff, Tagliamonte & Smith, 2005), a multivariate statistical analysis program for Windows. These factor groups were verb frequency (high/low), verb polarity (affirmative/negative), syntactic clause type (protasis, adverbial, adjectival, nominal) and dialogue/narration. Table 9 presents the results of the best run. The factor groups have been organized in descending order of magnitude (as determined by the range of the individual factor weights of each group). Individual factors within each group have also been organized from most to least favorable for selection of *-ra*. Those which have a factor weight greater than .5 (in bold) designate the contexts where *-ra* is preferred.

| Factor Constraints on *-ra* in the 17th Century | | |
|---|---|---|
| Dialogue/Narration | **Dialogue** **.72**<br>Narration .24 | Range: 48 |
| Syntactic Context | **Protasis** **.73**<br>**Adverbial** **.62**<br>Adjectival .47<br>Nominal .29 | Range: 44 |
| Verb Frequency | **High Frequency** **.66**<br>Low Frequency .41 | Range: 25 |
| Verb Polarity | **Negative** **.62**<br>Affirmative .47 | Range: 15 |
| Log likelihood = -253.833 | Significance = .042 | |

**Table 9: Multivariate Analysis Results for the 17th Century**

All four factor groups were selected as significant in the selection of *-ra*. A clear hierarchy of constraints emerges when these groups are compared. As determined by the ranges of each factor group, dialogue/narration was selected as the most influential factor group, closely followed by syntactic context, then verb frequency and, finally, verb polarity. Note that dialogue/narration and syntactic context also have the greatest magnitudes of impact since their ranges nearly double that of verb frequency and triple that of verb polarity. As such, the extent of *-ra*'s grammaticalization in the 17th century may be described.

### 5.4.1 Dialogue/Narration

The strong favoring of *-ra* in 17th century dialogue suggests that it may have been generalizing in spoken language and thusly represented in the characters' speech patterns. It is imperative, however, to keep in mind that dialogue in literature is not a direct representation of spoken language: it is crafted by a single author who has his own idiolect (i.e. his own

preferences given his life experiences); it may or may not represent an explicit attempt to capture popular speech of the time; and it may not follow actual patterns of speech (e.g. the dramas were all composed in verse). Nevertheless, Poplack (2007) and Poplack and Malvar (2007), cited in Copple (2011), found that dramatic texts do tend to reflect conventions of spoken language. Moreover, in the 17th century, dramas, unlike novels, were not considered literature in its formal sense; rather, they were written for popular entertainment, thus increasing the chance for the language to be more colloquial (Ingber, 1998). Indeed, Lope de Vega, for example, was known to try to reflect popular speech in his dramas (Parker, 1998). The dialogue of a novel could also assumingly attempt to reflect popular speech, but since novels were mostly accessible to the educated elite, it is possible that conventions of more "proper" language would characterize them (in this case, disfavoring *-ra*). Even if this were so, the text's author, regardless of genre, might attempt to intentionally replicate the tendencies of spoken language to make characters more realistic and thus relatable to their audience[12]. If capturing innovative speech patterns were a true concern of each author, then it would be expected that all dialogue be more innovative than narrative. Indeed, relative frequencies of *-ra* reveal that dialogue was a favorable context, regardless of genre. Although dialogue in novels does not demonstrate majority usage of *-ra* (N = 34/77, 44%), it is certainly more favorable for *-ra* than narration (N = 40/188, 17.5%), offering further evidence that *-ra* was most likely gaining ground in spoken language in the 17th century.

[12] This is where analyzing patterns of use through the lens of each character's sociolinguistic profile would be interesting. Favoring of *-ra* by certain social groups might provide evidence of where in society the change originated and first generalized.

### 5.4.2 Syntactic Context

Syntactic context, with a range of 44, was selected as the second most influential factor group for the 17th century. Within the factor group, a clear hierarchy of clause type also emerges. As shown in Table 9 above, protases were the most favorable for use of *-ra*, followed by other adverbial clauses. Adjectival and nominal clauses were not selected as favorable (they both have a factor weight of less than .5); although, *-ra* has encroached on adjectival clauses. Favorable use in protases is logical given their history in the early development of Imperfect Subjunctive *-ra*, but why might *-ra* generalize next in other adverbial clauses? If we return to the idea of clause interdependence and its relationship with grammaticalizing forms, a very plausible explanation may be provided.

#### 5.4.2.1 Degree of Clause Dependence and Grammaticalization

If we compare the hierarchy of clause types produced by the variable-rule analysis with the hierarchy of dependent clauses discussed in section 4.2.2, we find that it is an exact match. Thus, it would appear that *-ra*'s generalization was connected to its growing association as a subjunctive form in increasingly embedded subordinate contexts. This finding aligns with other observations of grammaticalizing forms (Bybee *et al.,* 1994). As noted above, like other grammaticalizing indicatives becoming subjunctives, *-ra* developed its hypothetical value and generalized first in main clauses since periphrastic *haber + past participle* had taken over the pluperfect indicative function in those same clause types (Klein-Andreu, 1991). Additionally, Bybee *et al.* (1994) theorize that subordinate uses of grammaticalizing forms are among the last to develop on a cline, a position which is corroborated by the observed spread of *-ra* across the clause types to increasingly syntactically dependent clauses presented above. As *-ra* expanded into dependent clauses, it appeared first in conditional protases before generalizing to adverbial

clauses, which are relatively independent yet still tend to rely upon another clause for full meaning. The fact that protases were selected as the most favorable for *-ra* corroborates what many (Klein-Andreu, 1991; Ward, 1991; Penny, 2002; Pitloun, 2006) have noted; this is, the protasis served as the entry point for subjunctive *-ra* as it made its way into subordinate contexts. The transition from main clause to dependent clause via the protasis would have been a logical transition for *-ra* for both syntactic and semantic reasons.

Syntactically, as was previously mentioned, protases and apodoses generally cooccur, thus allowing them to become a conventionalized structure. Usage-based approaches to language suggest that these types of constructions can develop independently since their frequency and pragmatic associations set them apart from other, less-conventionalized structures (Bybee, 2006). Also recall that it is not uncommon for the verb form used in the apodosis to also be employed in the protasis, even when this is not the prescriptive combination (cf. Lavandera [1984] found use of the conditional in both clauses and Silva-Corvalán [1985] found use of the imperfect indicative in both clauses, cited in Montrul [2018]). Higher frequency of *-ra* in conditional apodoses could facilitate its jump to the protases as, historically, these were both contexts for *-se* as well (Penny, 2002). Simultaneously, protases and apodoses work in tandem semantically to express equally hypothetical ideas. It is true that the protasis conveys a condition and the apodosis a result, but neither was/has been realized which may help explain the tendency toward formal assimilation of the two clauses. Be this as it may, recall that Klein-Andreu (1991) argued that formal assimilation between the two clauses, which allowed *-ra* to be used as a subjunctive in a subordinate clause for the first time, was not enough to initiate its spread to other dependent contexts. Rather, she proposed that indicative *-ra*'s relegation to pragmatically low-focus, and thus non-assertive, subordinate clauses is what allowed it to transition from main clause

subjunctive to subordinate subjunctive contexts, beginning with other adverbial clauses via the protasis.

In the 17th century, *-ra* was still disfavored in adjectival contexts, which follow adverbials in the dependence hierarchy. The adjectival clauses in the corpus tend to be restrictive, which are more dependent than appositives. In (35), although the restrictive clause could be removed and a complete thought would remain, it would take away from the description of the unspecified item to be purchased.

(35) …y comprar por su dinero alguna cosa que **pudiese** decir que la había hurtado

'…and buy with his money something that he **could** say he had stolen'

GIT. 17th cen.

However, as adjectival clauses only slightly disfavor *-ra,* its progressing generalization is still evident.

Finally, as the most embedded of the dependent clause types, it follows that *-ra* would appear in nominal clauses least as this represents the context for the most advanced stage of *-ra*'s grammaticalization and the context where use of *-se* would be conserved. In this way, *-ra*'s generalization in the other, more assertive clause types would have relegated *-se* to the most subordinate clauses much like *haber + past participle* constrained indicative *-ra* to these same contexts. Of interest, then, may be to examine which nominal clauses first began to accept *-ra,* leading to its eventual generalization.

Parallel structures and priming have been shown to influence variant selection of nearby analogous forms (Torres Cacoullos, 1999; Rosemeyer & Schwenter, 2019). In relation to selection of *-ra*, it would be expected for nominal clauses headed by a *-ra* verb, such as *quisiera que* 'I would like that', to be followed by another *-ra* form. Branza & van Heuven (2005)

examined this phenomenon and found an increased usage of *-ra* in clauses governed by a *-ra* verb (p.32). In the present corpus, a similar effect was found. In the 17th century, of the 20 nominal clauses governed by a verb conjugated with *-ra* (e.g. *quisiera, dijera, plugiera,* etc.), 19 were followed by a *-ra* form. This priming effect could have allowed *-ra* to begin to generalize in this context as its use in these specific clauses became routinized. Indeed, evidence of its spread is found in that these 19 clauses do not represent even half of the total nominal occurrences of *-ra* (N = 19/49 or 39%) in the 17th century corpus.

A growing shift of distributional patterns which increasingly favor more subordinate clauses alone most likely did not facilitate *-ra*'s spread through the clause types. The question then becomes what may have helped propel *-ra* from more to less independent contexts? The selection of verb frequency by the variable-rule analysis may very well help answer this question.

### 5.4.3 Verb Frequency

Frequency of form affects how language is used (Bybee, 2006; Torres Cacoullos, 1999). When high-frequency forms are perceived, they can come to serve as prototypes around which linguistic knowledge is organized. At times, they can be the most prone to linguistic change, going through processes of phonological reduction or semantic bleaching, or they can become entrenched, resisting, for example, processes of regularization (i.e. highly frequent irregular past-tense forms such as English *went* are not as likely to change to follow the regular pattern with *-ed*, in this case becoming *goed*). The selection of high-frequency verbs as favorable for *-ra* suggests that frequency also contributed to its generalization throughout subjunctive contexts.

In the 17th century, no frequency effect existed in main clauses. All verbs, regardless of frequency, almost categorically favored *-ra*. Although there are a few *-se* tokens among the high-

frequency verbs (N = 3/129, 2%), this can be considered a context exclusive to *-ra*[13]. Low-frequency verbs categorically appeared with *-ra* (N = 100/100, 100%). In contrast, dependent clauses exhibit a favoring effect in that high-frequency verbs were encoded with *-ra* (N = 118/181, 65%) at rates higher than the overall frequency of *-ra* in all dependent contexts (N = 236/521, 45%). If we compare the high-frequency verb types found in the 17th century independent clauses with those of the dependent clauses, a great deal of overlap is present[14]. This suggests that as they became conventionalized forms with *-ra* in independent-clause contexts, these same high-frequency forms could start appearing in dependent contexts by analogy. It has been shown that high-frequency forms, if they become conventionalized with an innovative variant, can facilitate linguistic change, especially if they develop a pragmatic function (Torres Cacoullos, 1999; Brown & Cortes Torres, 2012). So, if these highly frequent forms became conventionalized with *-ra* in main-clause contexts, as the data of this study suggests, and gained a stronger association as the form to be used for the Imperfect Subjunctive, then it would be a logical consequence that these same verb forms would begin to appear in novel contexts before the less frequent verbs. Indeed, if we observe the use of high- and low-frequency verbs according to clause type, presented in Table 10, we can see that *-ra* is more favored with high-frequency verbs across all clause types in comparison to low-frequency verbs, even in contexts in which *-ra* does not yet dominate.

[13] It is worth noting that one of the occurrences of *-se* was in the formulaic expression *fuese quien fuese* 'whoever it may be' and another in a conditional apodosis that followed a protasis with a *-se* verb.

[14] *Querer* 'to want', *haber* 'to have', *poder* 'to be able to' and *ser* 'to be' are shared between the two broad categorizations of syntactic clause types. *Saber* 'to know' and *dar* 'to give' also counted among the dependent clause high-frequency verbs; *saber* patterned with the other high-frequency verbs, showing a strong preference for *-ra,* while *dar* did not (N = 7/23, 30%).

| 17th Century Distribution of High- and Low-frequency Verbs by Clause Type | | | | | | |
|---|---|---|---|---|---|---|
| | ***-ra*** | | ***-se*** | | **Total** | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| **Protases** | | | | | | |
| High | 45 | **82** | 10 | **18** | 55 | **100** |
| Low | 43 | **74** | 15 | **26** | 58 | **100** |
| **Adverbial** | | | | | | |
| High | 31 | **74** | 11 | **26** | 42 | **100** |
| Low | 41 | **38** | 66 | **62** | 107 | **100** |
| **Adjectival** | | | | | | |
| High | 18 | **58** | 13 | **42** | 31 | **100** |
| Low | 9 | **27** | 24 | **73** | 33 | **100** |
| **Nominal** | | | | | | |
| High | 24 | **45** | 29 | **55** | 53 | **100** |
| Low | 25 | **18** | 117 | **82** | 142 | **100** |

**Table 10: Distribution of Verb Frequency by Clause Type (17th Century)**

Although these high frequency forms provided a gateway of sorts for *-ra* into the dependent clauses, the syntactic clause effect described above was still the stronger influence (recall that its magnitude of impact is nearly double that of frequency). Proof of the generalization of *-ra* by clause type is evident in that the protases favor *-ra* regardless of frequency and nominal clauses disfavor *-ra* regardless of frequency. In fact, the frequency effect

has already neutralized in the protases and is no longer statistically significant (*p-value*: 0.33). The frequency effect seen in the adverbial and adjectival clauses is worth noting. Here, the fact that the high-frequency verbs favor and the low-frequency verbs disfavor *-ra* suggests that these specialized verbs were the first to accept *-ra* in these dependent contexts, aiding the generalization across clause types. Indeed, here the frequency effect remains statistically significant[15].

### 5.4.4 Polarity

The final factor group selected was verb polarity (affirmative or negative). Negative contexts (presence of *no* or other negator before the verb encoded in *-ra*/*-se*) were statistically favorable for *-ra* in the 17th century. These results are somewhat surprising since negative contexts have generally been identified as linguistically conservative and less assertive. Bybee *et al.* (1994) have observed this cross-linguistically and Klein-Andreu (1991) considered negative contexts to be pragmatically "low focus". However, Pollán (2001) challenged this notion and argued that negative contexts may actually be more "high focus" than not. In her study of perfective past indicative uses of *-ra* and *haber + past participle* alongside the Preterit in Galician Spanish, she found that the non-preterit variants were favored in all contexts she had coded as "low focus" except for negative contexts, which she took as evidence of their more foregrounded status. She pointed out that providing a negative statement oftentimes serves to eliminate a possibility rather than provide an explanation of what actually occurred. Since the possibility being eliminated may have been the expected outcome, the negation would actually serve to foreground the message.

[15] *P-values* for the other clause types were as follows: Adverbials, 0.00; Adjectivals, 0.01; Nominals, 0.00.

If this position is adapted to the results of this study, then it is possible to suggest that *-ra* potentially had pragmatic value as an innovative form being used in a marked linguistic context. Negative contexts only accounted for 17% of the total *-ra*/*-se* occurrences. Consequently, this context may be considered marked, or less expected in discourse. Negating a verbal situation would, in theory, draw more attention to the message being presented since it carries a linguistic feature (in this case 'no') that is not expected by default. Simultaneously, as Copple (2011) pointed out by citing Haspelmath's (1999) "maxim of extravagance", using an innovative form in such contexts would further draw one's attention to the message being presented as it would also be more linguistically marked. This combination could pragmatically function as a way to make a message more salient and emphatic for its recipient. Like Pollán (2001) noted, even in the Imperfect Subjunctive, negation oftentimes serves to eliminate a specific possibility, as shown in (36).

(36) Mandaron los doctores que, por nueve días, no **hablase** nadie.

'The doctors ordered that, for nine days, no one talk.'

BUSC, 17th cen.

In this case, what the main character Pablo and his friend were allowed to do while recovering from being ill remains unanswered. This affirmation only serves to eliminate talking as one thing they may have done during their recovery period.

Semantically, the idea of negation may also correlate to *-ra*'s indicative past. In certain instances, not affirming a verbal situation denotes a negation of what actually occurred, as seen in (37).

(37) ¡Pluguiera a Dios **<u>no</u>** me **vieras**, o que allí fuera a tus ojos áspid, bruto, tigre o fiera!

'May it please God that you **had <u>not</u> seen** me, or that there I would have been an asp, brute, tiger or wild animal!

AHP, 17th cen.

Here, Estela is bemoaning the reality that the king did see her. This would contrast with the affirmative counterpart of this same situation, which would read as Estela's unrealized desire that he would have seen her. This ability to negate reality is especially true in conditional protases, whose communicative function is to hypothesize about what could be or could have been. If we again consider the distribution of *-ra* in negative contexts across clause types, we see that the negative protases were overwhelmingly favorable for *-ra* (N = 22/25, 88%)[16].

In sum, in the 17th century *-ra* was most likely the more frequent form in spoken language at the time. It was preferred in main clauses, protases, and other adverbial clauses and was emerging in adjectival clauses, and was most likely to appear with high-frequency verbs. Lastly, *-ra* possibly carried pragmatic weight as it was favored in negative contexts. Now let us turn to the 19th century to discuss how *-ra*'s constraints changed with time.

## 5.5 19th Century Multivariate Analysis

Recall that the relative frequency of *-ra* in the 19th century only increased by 4% when compared with the 17th century (see Table 8 above). This could simply indicate the slow nature of language change (Bybee *et al.*, 1994), but it could also indicate that *-ra* had become more prevalent in speech, allowing *-se* to gain a stronger association with written texts, even in dramas which are thought to much more closely replicate spoken language (Guzmán Naranjo, 2017;

[16]The relative frequency of negated *-ra* in the other syntactic contexts provides further evidence of *-ra*'s expansion through the clause types. Following the protases are the adverbials (N = 15/27, 55%), then adjectivals (N = 4/9, 44%) and finally nominal clauses (N = 6/27, 22%).

Copple, 2011). This latter reason seems more likely since, as previously mentioned, Romanticists were known to incorporate older linguistic conventions for stylistic effect. Although on the surface it does not appear that much change occurred between the 17th and the 19th centuries, this does not mean that processes of change were not still in motion.

In Table 11, below, are the results from the 19th century multivariate analysis alongside those from the 17th century for purposes of comparison. The same four factor groups were included for analysis (dialogue/narration, syntactic context, verb frequency and verb polarity). As in the 17th century, factor groups are ordered in descending order of impact (as determined by their ranges) and specific factors within each factor group that favor *-ra* (i.e. have a factor weight of .5 or greater) are in bold.

| **Comparison of 17th Century 19th Century Multivariate Analysis Rresults** | | | |
|---|---|---|---|
| **17th Century** | | **19th Century** | |
| Dialogue/ Narration | **Dialogue .72**<br>Narration .24<br>Range: 48 | Syntactic context | **Adjectival .66**<br>**Protases .56**<br>Adverbial .46<br>Nominal .36<br>Range: .30 |
| Syntactic Context | **Protasis .73**<br>**Adverbial .62**<br>Adjectival .47<br>Nominal .29<br>Range: 44 | Dialogue/ Narration | **Dialogue .58**<br>Narration .38<br>Range: .20 |
| Verb Frequency | **High .66**<br>Low .41<br>Range: 25 | Verb Frequency | **High .61**<br>Low .43<br>Range: 18 |
| Verb Polarity | **Negative .62**<br>Affirmative .47<br>Range: 15 | | |
| 19th century- Log likelihood = -196.910 Significance = 0.007 | | | |

**Table 11: 19th Century Multivariate Analysis Results (with 17th Century for Comparison)**

When compared with the 17th century, several key observations emerge. First, verb polarity no longer constrains the selection of *-ra*. The polarity effect has neutralized; affirmative contexts and negative contexts now accept *-ra* at rates that are no longer statistically significant. Second, syntactic clause type has replaced dialogue/narration as the most influential factor group, and, within this group, adjectival clauses have surfaced as the most favorable for *-ra*. Third, dialogue remains favorable, but its impact has diminished. Fourth, verb frequency, much like dialogue, still has a statistically significant impact, but less so than in the 17th century.

### 5.5.1 Pathways of Change

The results presented above help clarify the pathway of change that *-ra* followed as it generalized over time. The observed increase in relative frequency was not haphazard; *-ra* was gaining ground in particular contexts in the 19th century.

Beginning with syntactic context, the group which came to most strongly constrain *-ra* in the 19th century, we see evidence of *-ra*'s continuing expansion through the clause types. Adjectival clauses supplanted the protases and the adverbial clauses as the most favorable context for *-ra*, while nominal clauses remain the least favorable for *-ra*. The change in hierarchy may be slightly unexpected if we return to the notion of clause dependency (it would not be so surprising if adverbials overtook the protases and were still followed by adjectival clauses), but this may simply be due to the literary tendencies discussed in previous sections. The fact that adjectival clauses now favor *-ra* shows that *-ra* had continued to penetrate the clause dependency hierarchy to favorably appear in new contexts.

If we now turn to a breakdown of verb frequency by clause type, shown in Table 12, we will see more evidence of *-ra*'s continuing expansion through the clause types as having been led

by high-frequency verbs. Evidence of *-ra*'s continuing spread through the various degrees of clause embeddedness can be appreciated if we specifically focus on the nominal clauses.

| **19th Century Distribution of High- and Low-frequency Verbs by Clause Type** | | | | | | |
|---|---|---|---|---|---|---|
| | ***-ra*** | | ***-se*** | | **Total** | |
| | **N** | **%** | **N** | **%** | **N** | **%** |
| **Adjectival** | | | | | | |
| High | 23 | **71** | 9 | **29** | 32 | **100** |
| Low | 20 | **54** | 17 | **46** | 37 | **100** |
| **Protasis** | | | | | | |
| High | 19 | **61** | 12 | **39** | 31 | **100** |
| Low | 28 | **58** | 20 | **42** | 48 | **100** |
| **Adverbial** | | | | | | |
| High | 12 | **48** | 13 | **52** | 25 | **100** |
| Low | 15 | **42** | 21 | **58** | 36 | **100** |
| **Nominal** | | | | | | |
| High | 15 | **58** | 11 | **42** | 26 | **100** |
| Low | 20 | **28** | 52 | **72** | 72 | **100** |

**Table 12: Distribution of Verb Frequency by Clause Type (19th Century)**

While nominal clauses were still selected as disfavoring for *-ra* in the multivariate analysis, high-frequency verbs now exhibit a majority use of *-ra* (recall that in the 17th century this was not yet the case). In fact, nominal clauses are the only syntactic context in which high-

frequency verbs still have a statistically significant effect (*p-value* 0.01, significant) [17]. This contrasts with the 17th century in which all clause types except for the protases had a statistically significant frequency effect. From this, it can be said that by the 19th century usage of *-ra* had generalized among high- and low-frequency verbs except for in nominal clause contexts. Here, high-frequency verbs were still leading the way for *-ra* to progress into the final, most embedded of the subordinate syntactic contexts.

Dialogue remains favorable overall for *-ra*, although this effect has diminished in the 19th century. This supports the notion that spoken language tends to be more innovative than formal written texts (although dialogue is an indirect representation of spoken language). Indeed, even in the 20th century, statistically significant differences may be found between *-ra*/*-se* usage in speech and writing. Although this is a broad comparison, a Chi-square test for goodness of fit for *-ra*/*-se* occurrence in 20th century oral language and fiction literary texts based on the *Corpus de español* (Davies, 2002) found the difference to still be significant (*p-value* = 0, significant)[18]. Be this as it may, evidence of *-ra*'s spread to new contexts may still be appreciated if we again turn to narration from the novels. In the 19th century, narration, while still not showing majority use of *-ra* (N = 48/117, 41%), exhibits higher preference for *-ra* than it did in the 17th century (N = 40/228, 18%).

[17] *P-values* for the other clause types were as follows: Adjectivals, 0.13; Protases, 0.79; Adverbials, 0.62, all not significant.

[18] 20th century counts of *-ra* in fiction N = 16,684/18,954, 88%; oral N = 7,079/7,770, 91%

# Chapter 6 - Conclusions and Directions for Future Research

## 6.0 Summary of Findings

This study examined variation in the use of *-ra*/*-se* in the 17th and 19th centuries in order to determine how use of each variant changed over time. In the 17th century, there is evidence to suggest that *-ra* was the more popular morpheme in everyday speech since it was favored in dialogue. It can also be concluded that *-ra* was still developing its association with dependent-clause uses of the subjunctive, slowly making its way to less independent contexts. It was the preferred form in main clauses, conventional conditional adverbials and other adverbials, while it was emerging in adjectival clauses and disfavored in nominal clauses. This aligns with other observations of grammaticalizing forms since subordinate uses represent the most advanced stage grammaticalization (Bybee *et al.*, 1994). The spreading of *-ra* across clause types may very well have also been propelled by the growing conventionalization of *-ra* with highly frequent verb forms. As these forms became commonplace in independent clauses, the routinization of the form, along with *-ra*'s accepted use in low-focus indicative contexts, allowed it to begin appearing in subordinate contexts that were increasingly more syntactically subordinate. Finally, *-ra* was likely perceived as a pragmatically innovative form since it was favored in marked negative contexts.

In the 19th century, processes of change were still in motion even though the overall relative frequency of *-ra*/*-se* usage only shifted slightly. Even though literary tendencies of the time may account for the slight overall change, the variable-rule analysis suggests that even in potentially choosing to employ *-ra*/*-se* at more balanced rates for stylistic purposes, the authors' usage patterns still reveal that change was continuing to occur along the paths first observed in the 17th century. That is, *-ra* was expanding in increasingly subordinate syntactic contexts and

frequency effects, although diminishing, were still found. The grammaticalization of *-ra* was not complete since it was still disfavored in nominal clauses, the most subordinate. Nevertheless, evidence of *-ra*'s eventual generalization in this context may be seen in the significant frequency effect still at work in this context. The selection of dialogue over narration suggests the continued association of *-ra* with spoken language, even amidst writers who may have intentionally utilized *-se* in this context.

## 6.1 Limitations and Directions for Future Research

Using literature, even dramas, to analyze language change does present certain limitations. Although writers may attempt to replicate usage patterns of their time, literature will never present a direct representation of spoken language (the ideal source of linguistic data) since it is sensitive to author stylistic choices. However, since samples of spoken language are not available for historical research, we must rely on written documents for insight into the evolving patterns of use of linguistic forms over time. Perhaps a similar analysis of corpora of other genres would further illuminate the path *-ra* followed to replace *-se*.

Given that narration and dialogue pattern differently and that certain factor groups, such as grammatical person, had to be excluded from the analysis due to their interaction with them, further work could examine dialogue or narration only. Focusing on one text type would help to determine if these factor groups played a role in *-ra*'s evolution. Furthermore, an analysis of a corpus of texts written prior to the 17$^{th}$ century could trace the development of *-ra* as a developing subjunctive morpheme in main clauses. Of interest could also be the development of routinized expressions such as *(yo) quisiera* 'I would like', *plugiera a Dios* 'May it please God' and *pudiera* 's/he could/it could be' and the role they may have had in the transition of *-ra* from an indicative to a subjunctive. Additionally, this corpus or a similar one could be re-coded for

factors relating to pragmatic focus, such as those utilized by Klein-Andreu (1991) or Pollán (2001) to see if *-ra* was favored in foregrounded contexts over *-se*. Finally, although *-se*'s use is rare, corpora of modern usage could be analyzed to identify where specifically *-se* may still be used. If *-se* is to *-ra* what indicative *-ra* was to *haber* + *past participle*, then it would be expected that *-se* be limited to particular subordinate contexts in modern usage.

## 6.2 Concluding Thoughts

Language change is a complex process that ought to be subjected to detailed analysis so as to be fully understood. Surface distributions captured by relative frequencies do not tell the whole story and may even lead one to overlook the fact that significant change has occurred (recall that the increase in relative frequency of *-ra* between the 17$^{th}$ and 19$^{th}$ centuries was not statistically significant). By further investigating particular contexts of use, as was done in this study, the underlying differences between competing forms may be teased apart, allowing clearer pictures to be drawn and better stories to be told.

# Bibliography


Alcina Franch, J. (1969). Estudio preliminar in. J. Alcina Franch (Ed.), *Novelas ejemplares de Miguel de Cervantes* (pp. 11-28). New York, NY: Three Lions Incorporated.

Azevedo, M. M. (2009). *Introducción a la lingüística española.* Upper Saddle River, NJ: Prentice Hall.

Branza, M., & van Heuven, V. J. (2005). Linguistic variation in the subjuntivo imperfecto in Spanish America in the 16th century. *Linguistics in the Netherlands, 22*, 25-36.

Butt, J., & Benjamin, C. (2011). *A new reference grammar of modern Spanish* (5th ed.). London: Hodder Education.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*(4), 711-733.

Bybee, J., Perkins, R. & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world.* Chicago, IL: The University of Chicago Press

Copple, M. T. (2009). A diachronic study of the Spanish perfect(ive): Tracking the constraints on a grammaticalizing construction, ProQuest Dissertations and Theses.

Copple, M. T. (2011). Tracking the constraints on a grammaticalizing perfect(ive). *Language Variation and Change*, *23*(2), 163-191.

Copple, M. T. (2014). Following the path: An emerging perfect(ive) through temporal reference. In J. Bres, & E. Labeau (Eds.), *Evolution in Romance verbal systems* (pp. 55-90). New York, NY: Peter Lang

Davies, M. (2002). Corpus del español (100 million words, 1200s1900s). Available online at http://www.corpusdelespanol.org

DeMello, G. (1993). -Ra vs. -se subjunctive: A new look at an old topic, *Hispania*, *76(*2), 235–244.

Guzmán Naranjo, M. (2017). The se-ra alternation in Spanish subjunctive. *Corpus Linguistics and Linguistic Theory*, *13*(1), 97-134.

Hanna, C. M. (2012). Estudio diacrónico preliminar de la variación -se/-ra en España, *Plaza: Dialogues in Language and Literature*, 2.2. Available online at: https://journals.tdl.org/plaza/ index.php/plaza/article/view/6350/pdf_431

Hopper, J. & Closs Traugott, E. (2006). *Grammaticalization* (2nd ed.). New York, NY: Cambridge University Press

Hualde, J. I., Olarrea, A., Escobar, A. M., & Travis, C. E. (2010). *Introducción a la lingüística hisánica* (2nd ed.). New York, NY: Cambridge University Press.

Ingber, A. (1998). Lope Félix de Vega Carpio. In Parker, M. (Ed.). *Spanish dramatists of the Golden Age : A bio-bibliographical sourcebook*. (pp. 229-243) Westport, CT: Greenwood Press.

Kempas, I. (2011). On variation in the frame of free choice between cantara and cantase in Peninsular Spanish. *Moenia, 17*, 243-264.

King, E. (1962). What Is Spanish Romanticism? *Studies in Romanticism*, *2*(1), 1-11.

Klein-Andreu, F. (1991). "Losing ground: A discourse-pragmatic solution to the history of *-ra* in Spanish". In S. Fleischman & L. R. Waugh (Eds.), *Discourse-pragmatics and the verb: Evidence from romance* (pp. 164-178). London and New York, NY: Routledge,

Labov, W. (1986). The social stratification of (r) in New York City department stores. In *Dialect and Language Variation* (pp. 304-329).

Lapesa, R. (1981). *Historia de la lengua española* (9th ed.). Madrid: Gredos.

López García, A. (1999) Relaciones paratácticas e hipotácticas in V. Demonte & I. Bosque (Eds.), *Gramática descriptiva de la lengua española* (pp. 3509-3547). Madrid: Espasa.

Lunn, P. (1989). Spanish mood and the prototype of assertability. *Linguistics, 27*(4), 687-702.

Lunn, P. (1995). The evaluative function of the Spanish subjunctive In S. Fleischman & J.L. Bybee (Eds). *Modality in grammar and discourse.* (pp. 429-449). Philadelphia, PA: J. Benjamins.

Luquet, G. (2004). The subjunctive imperfect tense and the restructuring of the Spanish verb system towards the end of the Spanish Golden Age. *Studia Romanica Posnaniensia, 31,* 361-368.

Martínez, G. A. (2001). Linguistic politics and social contact in Mexico-Tejano Spanish: The -ra vs -se opposition in Texas during the nineteenth century. *Hispania, 84*(1), 114-124.

McKinnon, S. (2018). An initial examination of imperfect subjunctive variation in Catalonian Spanish: A contact linguistics and usage-based approach. In J. E. MacDonald (Ed.), Contemporary trends in Hispanic and Lusophone linguistics: Selected papers from the Hispanic Linguistic Symposium 2015 (pp. 333-353) Philadelphia, PA: John Benjamins.

Montrul, S. (2018) Morphology, syntax, and semantics in Spanish as a Heritage Language in K. Potowski (Ed.), *The Routledge Handbook of Spanish as a Heritage Language* (pp. 145-162). Routledge.

Parker, M. (1998). *Spanish dramatists of the Golden Age: A bio-bibliographical sourcebook.* Westport, CT.: Greenwood Press.

Penny, R. (2002). *A History of the Spanish Language*. (2nd ed.). New York, NY: Cambridge University Press.

Penny, R. (2004) *Variation and change in Spanish.* New York, NY: Cambridge University Press

Pharies, D. (2015). *Breve historia de la lengua española* (2nd ed.). Chicago, IL: The University of Chicago Press

Pitloun, P. (2006). El uso del imperfecto de subjuntivo en el habla culta de Costa Rica. *Revista de Filología Española*, *86*(2), 339-371.

Pollán, C. (2001). The expression of pragmatic values by means of verbal morphology: A variationist study. *Language Variation and Change*, *13*(1), 59-89.

Ramírez Luengo, J. L. (2001). Alternation of the forms -ra/-se in nineteenth-century Uruguayan Spanish. *Estudios Filológicos*, *36*, 173-186.

Ridruejo, E. (1999) *Modo y modalidad: El modo en las subordinadas sustantivas* in V. Demonte & I. Bosque (Eds.), *Gramática descriptiva de la lengua española* (pp. 3211-3251). Madrid: Espasa.

Rosemeyer, M., & Schwenter, S. (2019). Entrenchment and persistence in language change: The Spanish past subjunctive. *Corpus Linguistics and Linguistic Theory*, *15*(1), 167-204.

Sankoff, D., Tagliamonte, S. A., & Smith, E. (2005). GOLDVARB X: A multivariate analysis application for Macintosh and Windows. Department of Linguistics, University of Toronto, and Department of Mathematics, University of Ottawa. Available at: http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm.

Serrano, M. J. (1996). The subjunctives -ra and -se in conditional sentences. *Estudios Filológicos*, *31*, 129-140.

Silva-Corvalán, C. (2001) *Sociolingüística y pragmática del español.* Washington D.C. Georgetown University Press

Sorenson, T. (2016). The use of -ra and -se past subjunctive forms in educated written Spanish. *Bulletin of Hispanic Studies*, *93*(8), 821-841

Tagliamonte, S. A. (2006) *Analysing sociolinguistic variation*. New York, NY: Cambridge University Press

Torres Cacoullos, R. (1999). Construction Frequency and Reductive Change: Diachronic and Register Variation in Spanish Clitic Climbing. *Language Variation and Change, 11*(2), 143-70.

Torres Cacoullos, R. (2001). From lexical to grammatical to social meaning. *Language in Society, 30*(3), 443-478.

Veiga, A. (1999). Are there really conditionals in *si tuviera* or subjunctive uses of *cantara* in the poema de Fernan González? *Moenia, 5*, 271-306.

Ward, A. (1998). The -ra verb form in the conditional sentences of the *Estoria de Espanna*. *Verba, 25*, 127-141.

Williams, A. (1982). The use of the -ra and -se forms of the past subjunctive in Navarre. *Hispania, 65*(1), 89-93.

Wolford, T. E. (2005). Sociolinguistic conditioning, variation, and change in Spanish past subjunctive morphology and hypothetical discourse in the Southwest from the colonial period to the early 20th century. Available from ProQuest Dissertations & Theses Global.

# Appendix A - Codes for In-text Examples

## 17th Century

AC- *El amor constante*, Guillén de Castro

AHP- *Amor, honor, poder,* Calderón de la Barca

BUSC- *El buscón*, Quevedo

GIT- *La gitanilla*, Cervantes

VDM- *La vengadora de las mujeres*, Lope de Vega

## 18th Century

GRAM- *Los gramáticos: Historia chinesca*: Juan Pablo Forner (1782)

## 19th Century

ADF- *El afán de figurar*, Carnerero

EA- *Escenas andaluzas*, Estébanez de Calderón

DA- *Don Álvaro*, Duque de Rivas

SDB- *El señor de Bembibre*, Gil y Carrasco

SDN- *El sí de las niñas*, Moratín

## 20th Century

TÍA- *Tía Julia y el escribidor*: Mario Vargas Llosa (1977)

DLC- *Debajo de la cama- Author unknown*

# Appendix B - Descriptions of Corpus Texts

## 17th Century Dramas

### *La vengadora de las mujeres 'The Avenger of Women'*

Written by the innovative playwright Félix Lope de Vega (Ingber, 1998), this drama features a female protagonist, the Infanta Laura. Much to the disdain of her brother, the king, Laura refuses to get married and declares herself the "avenger of women" after having grown frustrated with how men and women are portrayed in the various books she has read. Laura rejects all who pursue her outright but allows Lisardo, a potential suitor in disguise, to be her secretary. To humor her brother, Laura allows a jousting tournament and writing competition to be arranged, the victor of which is to be her husband. Laura disguises herself as a mysterious white knight and ends up winning the tournament. Having satisfied her need for vengeance against men through this, Laura professes her love for Lisardo and allows her ladies in waiting to marry her other suitors.

### *Amor, honor, poder 'Love, Honor, Might'*

Written by Pedro Calderón de la Barca, this drama centers around several complicated relationships that form between the royal family, their attendants and a family of nobles. Unbeknownst to him at first, Enrico, a nobleman, rescues the king's sister from an accident she suffered while out in the woods by his family's castle. This leads her, the king and their company to spend the night at his castle. Enrico and the Infanta fall in love, but through a series of miscommunications, neither openly admits their love for the other. Meanwhile, the king falls for Estela, Enrico's beautiful sister. He tries to use his power to get to her, but she outwits him. In a

later scene, Enrico injures one of the king's attendants in defense of his sister's honor, which lands him in prison. The Infanta, disguised as a man, tries to help Enrico escape, but he refuses, accepting the penalty for his actions against Teobaldo. In the end, Estela threatens to take her own life in protest of her brother's imprisonment and the king's abuse of power. At this, the king realizes and admits the error of his actions, which prompts Estela to agree to marry him, Enrico is freed and is allowed to marry the Infanta.

### *El amor constante 'Constant Love'*

In this drama written by Guillén de Castro, the king finds himself unhappy as he is in love with Nísida, the daughter of a duke, and not with his wife. Nísida, however, does not love him in return. Her heart still belongs to the king's brother, Celauro, who was imprisoned many years before. Celauro is released, and when he and Nisida have a moment alone, Nisida reveals that she had given birth to their son after his imprisonment and that she had given him a jewel-studded cross before sending him away with a servant, never to know the son's fate. Meanwhile, a young shepherd named Leonido, in a series of chance encounters, defends the Infanta and Celauro from various potentially fatal situations, gaining favor with both. Later, the king brings poison to Nisida, which she drinks. Before her death, she is able to receive her father's blessing for her union to Celauro. Celauro swears vengeance against the king, but ends up taking his own life. Prior to his death, he encounters Leonido, who he realizes is his long-lost son upon seeing the jewel-studded cross. Witnesses confirm the relationship and Leonido is crowned king. However, after receiving the crown, he passes it to the Infanta.

# 17th Century Novels

## *La gitanilla 'The Gypsy Girl'*

Miguel de Cervantes' most favored *novela* from his collection of *Novelas ejemplares* (Alcina Franch editor in Cervantes, 1969), *La gitanilla* tells the story of a young gypsy girl, named Preciosa, who is well known for her beauty and ability to sing and dance. On her way home from a social gathering with her adoptive grandmother, a wealthy young man, don Juan de Cárcamo, professes his love for Preciosa and his intentions to marry her. Preciosa agrees but under the condition that he should prove his devotion to her by becoming a gypsy. Although he was not expecting this, he agrees, and under the new name Andrés Caballero, he is accepted into her community. As they are traveling, they stop in a small town where a young lady falls madly in love with Andrés. Upset that he rejects her advances, she plants some of her jewels in his belongings before they leave and accuses him of stealing them. During the ensuing confrontation, Andrés, to defend his own honor, mortally wounds a soldier and is arrested. When the town *corregidor* and his wife come to assess the situation, a strange turn of events occurs. The wife reacts strongly to Preciosa´s presence, as if something deep inside her recognized the young girl, since she is of the same age as her own lost daughter. Preciosa´s grandmother, prompted by the *corregidora´*s comments, retrieves a small jewelry box. In confidence, she reveals to the woman that Preciosa is in fact her daughter and that she had taken her years before. Surprisingly, the couple is so thankful to be reunited with their daughter that they do not punish the gypsy woman for her transgression. In the end, Andrés is released from prison and all come together to celebrate his marriage to Preciosa.

### *El buscón ´The Swindler´ or Historia de la vida del Buscón, llamado Don Pablos, ejemplo de vagamundos y espejo de tacaños ´History of the Life of the Swindler, Called Don Pablos, Model of Vagabonds and Mirror of the Stingy´ (Books I and II)*

Francisco de Quevedo's *novela picaresca* tells the story of a young boy's (Pablos) misadventures. Born to a couple who makes a humble living as barbers and thieves, Pablos sets out to make a better life for himself. However, in each new opportunity (schools, apprenticeships, etc.), he either ends up getting in trouble or faces abuses which lead to his removal from each place. In true *pícaro* fashion, Pablos learns to get by as he can, conniving and stealing to meet his basic needs. Upon receiving word of his father's death, he travels to receive his inheritance from his uncle, meeting a series of unique individuals along the way. Finding life with his uncle to be unbearable, he escapes with his inheritance to make his way to Madrid. On this journey, he meets other characters, such as an *hidalgo*, who teach him other street smarts to get by in various social situations and make his way into the court in Madrid.

## 19th Century Dramatic Texts

### *El sí de las niñas ´The Maiden´s Consent´*

Written by Leandro Fernández de Moratín, this drama takes place in an inn and occurs over the course of one night. Doña Irene has come to retrieve her daughter, Doña Francisca (Paquita) from a local convent where she had been staying with her nun aunt. Paquita is engaged to Don Diego, nearly 45 years her senior, but Paquita is in love with Carlos, Don Diego´s nephew. Carlos, under the pseudonym of Don Felix, had met and fallen in love with Paquita a year prior, but had never let it be known to his uncle. Upon receiving Paquita´s plea for help, Carlos travels to the inn where he discovers that the man pledged to Paquita is none other than his uncle. After a confrontation with his nephew over the matter, Don Diego has a change of

heart and convinces Paquita´s mother, to the surprise of all, to permit the marriage of her daughter and his nephew.

### *Don Álvaro o la fuerza del sino ´Don Álvaro or the Force of Destiny´*

By Ángel de Saavedra, better known as the Duque de Rivas, this drama tells the story of the repercussions of an unfortunate event that ended a forbidden love affair. The mysterious Don Álvaro has arrived in Sevilla and falls for Leonor, the marques of Calatrava's daughter. However, her father prohibits the union. Álvaro and Leonor plot to elope, but on the night of their escape, they are caught. Although he surrenders his pistol, it accidently goes off when it's tossed to the ground, mortally wounding the marques. It is later revealed that, following the events of that fateful night, Álvaro and Leonor part ways and flee. Álvaro is believed to have set sail for the Americas, and Leonor ends up seeking refuge in a local convent since her brothers are searching for her and Álvaro to avenge their father's death. Meanwhile, Álvaro, who has become a respected member of the military, has a chance encounter with Leonor's brother Don Carlos in Italy. He saves Carlos's life after a swordfight, which causes them to become good friends, but both conceal their true identity. Álvaro is wounded in battle, and it is feared that he will lose his life. As his last wish, he instructs Carlos to burn a parcel of letters in his suitcase. Upon going to do so, Carlos discovers that his new friend is actually his enemy when he sees a picture of his sister with the letters. Things are further complicated when Álvaro pulls through surgery and survives his wound. Upon regaining his strength, Carlos confronts him, the two dual, and Álvaro wins. Álvaro returns to Spain and seeks refuge disguised as a monk in the same convent where Leonor is, unbeknownst to him. Leonor's other brother, Don Alfonso, manages to find him there and also challenges him to dual. After wounding Alfonso, Álvaro seeks help from

the mysterious hermit, Leonor. She rushes to the aid of her brother, but he stabs her. Both siblings die, and Álvaro runs off into the mountains, leaving the other monks in shock.

### *El afán de figurar ´The Thirst for Appearances´*

Written by Juan de Carnerero, this drama, set in Paris, revolves around the relationships and ambitions of Derval, an aspiring writer, and his close friend, the baron of Monsernin. The baron is known for wavering, which becomes a source of drama as he has rescinded his desire for his sister, Emilia, to marry Derval. Even so, Derval entrusts the baron with the one remaining copy of a controversial publication he had written. The baron's almost fiancé, the Countess of Monreal, takes possession of the pamphlet to submit to the authorities with the intent of ruining Derval's chance of receiving a top governmental position for which the baron is also contending. Although the baron is uneasy about how this will affect his long-time friendship, his feelings for the countess and his own ambition get in the way of him stopping her. In the end, the conniving countess's plan backfires, and Derval gets the promotion due to the fact that the Minister who received his writing thought very highly of his ideas. The baron blesses the union of Emilia and Derval and leaves the countess to return to his country house.

## 19th Century Novels

### *El Senor de Bembibre ´The Gentleman from Bembibre´ (Chapters 1-10)*

Written by Enrique Gil y Carrasco, this dramatic novel tells the tale of Don Álvaro, also known as the Señor de Bembibre, and Beatriz, the only remaining heir to her father´s estate, who are madly in love. However, due to her family´s circumstances, Beatriz´s parents are planning to arrange for her to be married to the even more wealthy Count of Lemus. Beatriz refuses, threatening to become a nun before marrying a man whom she does not love. To teach her a

lesson, her father sends her to the local convent. While she is there, her father visits to inform her that her marriage to the count has been definitively arranged and will occur in three days. Although her protest is futile, she boldly refuses her father's wishes. She then plots with her maidservant to go seek help from Álvaro. At the close of Chapter 10, where data extraction stopped, Álvaro and the handmaiden are beginning to put their rescue mission into motion.

### *Escenas andaluzas ´Scenes from Andalucía´ (Pages 1-48 of the PDF version available online through the Biblioteca Virtual Universal*[19]*)*

Serefín Estébanez Calderón's collection of stories recounts isolated encounters between locals who frequent neighborhood taverns and the lives of other prominent figures. Conversations about unimportant topics, such as the origin of the bolero dance and love interests, were retold. Also, the life of one Manuel Gázquez, a blacksmith, was recounted for many of the pages of the section used for extraction.

[19] The text may be accessed here: https://www.biblioteca.org.ar/libros/70209.pdf