Master's Theses                                      Master's Theses and Graduate Research

Summer 2017

# Quantifying the Functional Consequences of Spanish [s] Lenition: Plural Marking and Derived Homophony in Western Andalusian and Castilian

Mary Catherine Moran Ryan
*San Jose State University*

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

QUANTIFYING THE FUNCTIONAL CONSEQUENCES OF SPANISH [S]
LENITION:  PLURAL MARKING AND DERIVED HOMOPHONY
IN WESTERN ANDALUSIAN AND CASTILIAN


A Thesis

Presented to

The Faculty of the Department of Linguistics and Language Development

San José State University



In Partial Fulfillment

of the Requirements for the Degree

Master of Arts



by

Mary Moran Ryan

August 2017

The Designated Thesis Committee Approves the Thesis Titled

QUANTIFYING THE FUNCTIONAL CONSEQUENCES OF SPANISH [S]
LENITION:  PLURAL MARKING AND DERIVED HOMOPHONY
IN WESTERN ANDALUSIAN AND CASTILIAN

by

Mary Moran Ryan

APPROVED FOR THE DEPARTMENT OF LINGUISTICS AND
LANGUAGE DEVELOPMENT

SAN JOSÉ STATE UNIVERSITY

August 2017

| | |
|---|---|
| Dr. Daniel Silverman | Department of Linguistics and Language Development |
| Dr. Hahn Koo | Department of Linguistics and Language Development |
| Dr. Chris Donlay | Department of Linguistics and Language Development |

ABSTRACT

QUANTIFYING THE FUNCTIONAL CONSEQUENCES OF SPANISH [S]
LENITION:  PLURAL MARKING AND DERIVED HOMOPHONY IN WESTERN
ANDALUSIAN AND CASTILIAN

By Mary Moran Ryan

In this thesis, a new methodology is proposed for investigating Spanish [s] lenition

(sound weakening or loss) via morphological analysis instead of phonetics. Word-final

[s] is a morphological plural marker in Castilian Spanish, but is rarely produced in

Western Andalusian Spanish (WAS). It is often asserted in the literature that the loss of

[s] in WAS requires plurality to be expressed via alternative means. The results of this

study rule out lexical and morpho-syntactic compensation for [s] lenition in WAS in

several previously untested domains, and imply that there is no functional motivation in

Modern Spanish driving a need for compensation for word-final [s] lenition on nouns or

determiners. This investigation is built on a predictable calculation of the environments in

which the loss of [s] may result in derived singular/plural homophony in WAS nouns.

This is used to quantify potential semantic ambiguity. A frequency comparison of 27,366

WAS and Castilian nouns, across 60 specific Determiner + Noun phrase environments,

finds no significant differences between the dialects in the type or token frequencies of

numerically ambiguous nouns, nor in 98.7% of the tested phrase environments. When

taken in context with studies excluding phonetic compensation in WAS, the current

results suggest that the low semantic relevance of word-final [s] in Modern Spanish is a

potentially far-reaching explanation for the variable manifestations of [s] lenition

experienced in Spanish dialects across the world.

ACKNOWLEDGMENTS

# TABLE OF CONTENTS

**List of Tables**

# List of Abbreviations

| | |
|---|---|
| WAS | Western Andalusian Spanish |
| DET+N | Determiner + Noun |
| FEM INDEF ART | Feminine Indefinite Article: *una, unas* ("a / some / one") |
| [s] | A broad IPA transcription of the Spanish dental/alveolar fricative[1] |
| /s/ | [s] and its allophones (variants) |
| V-final | Vowel-final Noun |
| C-final | Consonant-final Noun |
| Ø | Zero Marker |
| *Italics* | Orthographic Spanish Transcription |
| < > | Graphemes |
| " " | English gloss |

---

[1] This thesis attempts to discuss the specifics of Spanish phonetics and phonology in as theory-neutral a manner as possible. With this goal in mind, square brackets [ ] are purposefully used in many places where virgules / / might traditionally have been employed. Both symbols retain their traditional meanings. This approach is intended to allow the analysis to remain focused on physical sound wherever practical, rather than the more abstract ideas of phonemes and allophones.

## 1. Introduction

Speakers of the Castilian dialect of Spanish distinguish between [la nota] "the note" as singular, and [las notas] "the notes" as plural, with word-final [s] functioning as a morphological plural marker. Not every Spanish dialect uses [s] in this way. Speakers of Western Andalusian Spanish tend to use [la nota] as both a singular and a plural form.[2] This loss of singular/plural contrast is caused by a process of lenition (sound weakening or loss), which causes word-final [s] to neutralize toward zero in many Spanish dialects.

Word-final [s] lenition occurs in a wide array of patterns and acoustic variations across most Spanish speaking countries. The traditional description of how [s] lenition progresses is presented as [s] > [h] > ∅ (zero), with [h] representing not only the voiceless glottal fricative, but also sometimes standing in for aspiration or breathy phonation in the early literature. In this notation, in lenition dialects, [nota-s] might become [nota-h], and then may be variably or consistently produced as [nota]. For a few dialects, [nota] becomes the new plural form. This variable production does not result in function-negative consequences for speakers who share a dialect.[3] When [s] undergoes sufficient lenition that listeners can no longer distinguish between singular and plural forms, those forms may be considered derived homophones, like the singular [nota] and plural [nota] of Western Andalusian Spanish (WAS).

---

[2] WAS is distinct from the well-known dialect of Eastern Andalusian Spanish (EAS), which has some geographic overlap with WAS, and which is often asserted to use vowel allophony for plural marking. Lenition in EAS is beyond the scope of the current study.
[3] Speakers from differing dialects may require a period of acclimation before fluently understanding dialects with fewer phonemes (Labov, 1994).

The purpose of the current study is to test the often-repeated claim that [s] lenition results in a semantically critical loss of plural marking in Spanish, and therefore must trigger some method of compensation, in order to avoid communicative confusion. My hypothesis is that despite a demonstrable reduction of morphological plural marking, WAS speakers do not require compensation for the loss of word-final [s], because word-final [s] has a much lower semantic relevance in Modern Spanish than is generally acknowledged.

To test this hypothesis, I used the divergent, but predictable, plural morphologies of WAS and Castilian to identify which nouns and determiners in WAS retain structural number-cues without [s], and which do not. Those that do not are potentially semantically ambiguous in certain syntactic environments. If the proto-typical [s]-retaining dialect (Castilian), and the proto-typical [s] lenition dialect (WAS), can be shown to use morphologically-ambiguous nouns and determiners at the same rates, with no evidence of phonetic, lexical, or morpho-syntactic compensation, then this is strong evidence that no true semantic ambiguity is occurring, and no compensation is necessary in modern Spanish.[4]

This corpus comparison analyzes 27,366 nouns in WAS and Castilian Spanish, extracted in context, with their attendant determiners, from the freely available online transcripts of the PRESEEA-MA Corpus *(*Málaga Urban Vernacular, 2007) and the

---

[4] Though [s] lenition also occurs word-medially (which is lexical, rather than morphological), and additionally occurs on some verb inflections, these environments are beyond the scope of the current study. This investigation is focused exclusively on whether the lost phonetic contrast between singulars and plurals in WAS has semantic consequences.

PRESEEA Corpus (PRESEEA, 2014, selections from Madrid). I find that the usage of WAS nouns, and Determiner + Noun (DET+N) phrases, is comparable to the usage of nouns and phrases in the same environments in Castilian in 98.7% of instances, despite the 77% of WAS nouns that are vulnerable to derived homophony in the absence of word-final [s].

In the context of the combined results of Carlson 2012, Defior, Alegría, Titos, and Martos 2008, O'Neill 2005, and Ranson 1993, which effectively rule out contrastive phonetic compensation for [s] lenition in WAS, the results of the current study suggest that the semantic relevance of /s/ as a plural marker must be quite low in modern Spanish. I find no evidence of lexical compensation, and the only potential for morpho-syntactic compensation is found in phrases of the construction Feminine Indefinite Article + Noun (FEM INDEF ART + N), which are phrases with *una/unas ("a" or "some")*. This frequency difference is not seen in any other determiner environment.

The rest of this introduction includes a Background section (1.1) to put the major issues of Spanish [s] lenition in context, an overview of the current study (1.2), a primer on the relevant Spanish morphology (1.3), and an explanation of the questions that are asked and answered in this thesis (1.4). I also preview some of the study methodology in Section 1.2. The design of this study relies on concepts from modern and historical Spanish morphology, functional phonology, and computational linguistics. The study itself is straight-forward, but the methodology is more accessible if some key concepts are presented earlier rather than later.

## 1.1. Background

Spanish is the second most spoken language in the world, after Mandarin, and more than 20 countries claim it as a first language due to Spanish colonialism. Lenition is the weakening or loss of a sound, and in this case, it refers specifically to [s] losing saliency, typically by a reduction in the degree of constriction, hence a reduction in frication. The Castilian dialect is the historical dialect of Spanish government and literature, and is centered around Madrid, Spain. The [s] of Castilian is said to be apical in nature, compared to the [s] of Andalusia, which is described as a more laminal [s] (Romero, 1994).

WAS is spoken in southern Spain, and the inventory of sibilants is believed to have diverged from Castilian sometime between the 13th and 14th centuries, with [s] lenition established as a spoken norm in the Andalusia region by the end of the 16th century, (Penny, 2002). Most modern Spanish dialects have their origins in one of these two dialects, so a comparison between these two may provide information relevant to other dialects as well (Penny, 2002).

The question of how Spanish could have some dialects that mark plurals with morphological [s] and others that do not has been the subject of dozens of studies, most of which are focused on the possibility of phonetic compensation for the lost sound. Hernandez-Campoy and Trudgill (2002) provide a well-known list of 100 studies representing a cross-section of Spanish lenition research, primarily from the 1900s, before the wide availability of publicly accessible online data. Much of the more recent research has been focused either on word-medial [s] lenition, which is lexical rather than

morphological, or predominantly on acoustic production, documenting the rich acoustic

variation of /s/ production in Spanish (Carlson 2012, Erker, 2010; File-Muriel, 2010,

2012; File-Muriel & Brown, 2011; O'Neill 2005, 2009, Parrell, 2012; Ruch, 2013; Ruch

& Harrington, 2014; Torreira, 2007, 2012; Torreira, & Ernestus 2012).

   The assumption in the majority of previous studies, with Ranson 1993 as a notable

exception, is that [s] lenition dialects must have differences from non-lenition dialects

that allow them to signal the information previously carried by [s] in an alternative way.

The following three quotes illustrate the general opinion in the literature, and are included

here to illustrate a point, not to single out particular authors, as most lenition studies

include or at least imply such statements as well.

> In this position, the underlying /s/ is of great importance for the
> morphological system of Spanish due to the fact that the sibilant is the
> marker of plurality on the nominal and that of the second person singular
> on the verb. (O'Neill, 2005, p. 151)
>
> Spanish /s/ loss has often featured in discussions of functional
> compensation in language change for the good reason that the complete
> loss of postvocalic /s/ in word-final position in Spanish entails a loss of
> grammatical information and a large increase in ambiguity. (Hernandez-
> Campoy & Trudgill, 2002, p. 142)
>
> The object of this paper is to investigate the various factors constraining
> deletion versus retention of the plural marker, as well as the factors
> responsible for disambiguation in the case of marker deletion. The
> understanding of processes by which languages undergo lenition and
> deletion **of elements with a heavy functional load**, (emphasis mine) as
> well as the mechanisms by which they compensate for these deletions, is
> an issue of importance to general linguistic theory. (Poplack, 1980, p. 56)

   Statements like these, proposing that /s/ lenition causes a loss of grammatical

information, that /s/ is of great importance, and that /s/ has a high functional load in

Spanish,[5] are found throughout the literature, and with good reason. Plural nouns ending in [h], aspiration, breathy-phonation, lengthened vowels, and lowered vowels have been discussed in the literature for decades. Specific to WAS, breathy phonation is found word-finally on nouns produced with an interrogative intonation (O'Neill, 2005).

Additionally, there is ample cross-linguistic evidence that it is rare for a sound to undergo neutralization in circumstances that create homophones, such as the derived homophony created between singular and plural nouns in WAS. In an investigation of sound change in 153 languages, Gurevich (2004) finds that 92% of lenitions avoid phonological neutralization (neutralization that results in the loss of semantically relevant information). Silverman (2010), and Wedel, Kaplan, and Jackson (2013) also provide quantitative cross-linguistic evidence that sound alternations which propagate overwhelmingly tend to be those that do not have counter-functional consequences. In this context, the focus on compensation for Spanish [s] lenition appears to be supported by history, intuitive logic, and cross-linguistic data.

The fact remains, however, that no study thus far has been able to demonstrate that *listeners* can disambiguate isolated plural nouns from singulars based on phonetic cues alone. WAS listeners demonstrably cannot distinguish plurals from singulars when the nouns are isolated from external cues (Carlson, 2012; O'Neill, 2005), nor when they are placed in ambiguous carrier phrases (Carlson, 2012). They cannot tell the difference even when listening to recordings of their own voices reading the words and phrases. This is

---

[5] Functional load refers to the amount of "work," or contrast, for which a sound is responsible. This will be discussed in more detail presently.

consistent with the results of studies done in other [s] lenition dialects: Miller and Schmidt 2010 (Chile, Mexico), Poplack 1980 (Puerto Rico), and Terrell 1979 (Dominican Republic). If speakers are compensating for [s] lenition in WAS, it cannot be by way of phonetic replacement of [s].

Large-scale synchronic compensation for [s] lenition, phonetic or otherwise, is not the only possible answer for this sound change, however. The current study diverges from most lenition investigations to take a closer look at a proposal in Ranson 1993, which suggests that grammatical number marking with word-final [s] may be less semantically relevant than previously assumed, such that its loss may simply be tolerated, without the need for compensation in Modern Spanish.

The previous studies, many of which are discussed in more detail in Section 2, make it clear that WAS speakers do not need word-final /s/ to understand grammatical number, but there are still open questions about the possibilities of lexical or syntactic compensation. This study seeks to fill those gaps, and to strengthen the argument that [s] lenition is able to proceed based on its lack of semantic relevance.

**1.2. Overview of the Current Study**

One of the obstacles in trying to quantify the consequences of Spanish [s] lenition is that the number of ways this lenition can vary makes it a difficult variable to isolate. The use of word-final /s/ can differ greatly (even within dialects), by socioeconomics, education, gender, age, and register. There is also ongoing disagreement about how to categorize the various sounds attested at the end of some plural nouns in some dialects. As previously mentioned, the traditional description for the progression of lenition is

often presented as [s] > [h] > ∅ (zero), with [h] sometimes standing in as a "catch-all" symbol for aspiration or breathy-phonation. The issue, however, is that some studies code [h] and [s] together with the belief that either sound signals plurality, while others code [h] and zero together, because it has not been demonstrated that listeners can use [h], or other phonetic alternatives, to reliably identify plurality on out-of-context nouns. If the point is to know how speakers are signaling grammatical number, the differences in how data is collected and coded makes side-by-side comparison of these studies difficult.

Modern acoustic studies often avoid this notation completely. File-Muriel and Brown (2009) describe another difficulty with impressionistic coding, "Previous studies of s-weakening in Spanish have relied almost exclusively on the impressionistic coding of /s/. Not only is auditory transcription invariably influenced by the transcriber's background, but temporal and gradient acoustic details about the sound are concealed when tokens are represented symbolically." So in addition to the differing criteria studies may use to categorize the final sounds produced on nouns, there is also the problem that impressionistic coding captures whatever researchers or consultants are successfully understanding, but this gives no real information about whether they are gathering plural cues from context or from phonetics. Additionally, asking a listener for an assessment of how they are understanding plurality is fraught with difficulty, because it only reveals how that listener *believes* they are processing language, and not necessarily the actual process. For this reason, the studies that effectively isolate words or phrases in multiple ways, and multiple environments, and force active identification of number-marking, are the most reliable data currently available.

In WAS, the results of plural identification tests in O'Neill 2005, and Carlson 2012 suggest that regardless of what segmental or sub-segmental sounds speakers may produce word-finally on nouns, WAS listeners do not use that information contrastively to identify plurals. Nevertheless, an alternative way to study [s] lenition, that does not rely on acoustics, is replicable, and which can provide supporting evidence to previous studies, might be a useful addition to the field of Spanish [s] lenition.

With those criteria in mind, my study takes an approach to [s] lenition that has not been previously attempted. The methodology is designed around Spanish morphology instead of phonetics because Spanish plural morphology is highly predictable within each dialect, while Spanish phonetics are not. Using morphology avoids all of the previous questions of how to code word-final [s], [h], or zero. Structural number-cues that persist when [s] is lost can be calculated without reference to [s] or [h] at all, and instead, lexical and morpho-syntactic choices become the means of comparison.

Outside of the traditional plural morphology, grammatical number can be signaled by context, by verb morphology, or by historical structural differences on nouns or determiners that persist even when [s] is absent. As will be discussed in the next section, nouns that end in consonants, and masculine determiners, retain acoustic information that can be used to disambiguate plurals from singulars even in the absence of [s]. In this thesis, these historical differences are termed "cues" to grammatical number.

An additional benefit of using predictable morphological rules to investigate [s] lenition is that it allows working with larger data sets, because the rules can be applied via computational linguistic techniques. This methodology is not dialect specific, and

could be used to compare any Spanish dialect with any other. As previously mentioned, I chose to compare WAS and Castilian with the hope that a comparison of these two dialects would provide useful information relevant to other dialects as well. The idea to use derived homophony as a measure of potential semantic ambiguity was inspired by Silverman (2010, 2012).

## 1.3. Spanish Nominal Morphology

In Castilian, morphological plural marking on nouns consists of adding a word-final [s] if the noun ends in a vowel ( [nota] > [nota-s] ), and a word-final [es] if the noun ends in a consonant ( [profesoɾ] > [profesoɾ-es] ). Grammatical number in Spanish usually corresponds simply to "one" or "more than one," but this correspondence is not foolproof. Non-count nouns may be grammatically singular but semantically plural (like *la gente* "people"), and some nouns like *las gafas* (eye glasses) are semantically singular, but grammatically plural. These exceptions are limited however, and have no impact on the analysis because they still follow the respective morphological rules of each dialect.

Gender marking in Spanish is semantically motivated for animate nouns, but is arbitrary for inanimate nouns. While most nouns follow the paradigm of a final [o] for masculine, and [a] for feminine, there are also irregulars, like *el problema* "the problem" (MASC) or *la mano* "the hand" (FEM) in which the gender assignment is the reverse. For inanimate nouns like *nota*, gender is both arbitrary and fixed. Nouns that end in [e] or a consonant vary in gender assignment. Nouns ending in [i] or [u] may be of either gender, but are rare and almost exclusively loan words or truncations.

With the exception of a small number of loan words, every noun in Castilian carries overt number marking, because the consistent use of [s] or [es] on plurals means that the lack of those sounds indicates that the noun is singular. This does not hold true for lenition dialects. In any dialect, as soon as /s/ ([s] and any potential variants) can no longer be relied upon as a consistent plural marker, singular nouns are rendered structurally ambiguous if isolated from external cues. Table 1 documents the plural morphology for Spanish nouns in non-lenition and lenition dialects. The sample words are *amiga* ("friend," FEM), *amigo* ("friend," MASC), *mujer* ("woman"), and *profesor* ("professor").

Table 1

*Plural Cues on Castilian and WAS Nouns*

| | Vowel-final Noun | | Consonant-final Noun | |
|---|---|---|---|---|
| | Feminine | Masculine | Feminine | Masculine |
| **Non-Lenition** | | | | |
| Singular | [amiɣa] | [amiɣo] | [muxeɾ] | [profesoɾ] |
| Plural | [amiɣa-s] | [amiɣo-s] | [muxeɾ-es] | [profesoɾ-es] |
| Number Cues | 1 | 1 | 1 | 1 |
| **Lenition** | | | | |
| Singular | [amiɣa] | [amiɣo] | [muxeɾ] | [profesoɾ] |
| Plural | [amiɣa] | [amiɣo] | [muxeɾ-e] | [profesoɾ-e] |
| Number Cues | 0 | 0 | 1 | 1 |
| **Ambiguity potential** | Ambiguous | Ambiguous | Not ambiguous | Not ambiguous |

The singular forms are identical in non-lenition and lenition dialects, but in the plural forms of lenition dialects, word-final [s] is absent. Recall that it is immaterial whether [s]

is actually produced because the important measure here is about which nouns have the potential to become singular/plural homophones, and therefore ambiguous.

Specific to a full lenition dialect like WAS, for vowel-final (V-final) feminine and masculine nouns, the singular and plural forms are derived homophones and cannot be phonetically disambiguated without external cues. The V-final nouns have zero cues to grammatical number, but in a non-lenition dialect, like Castilian, they each have one cue. On the consonant-final (C-final) nouns, both the feminine and the masculine nouns retain the [e] of the previous [es] morphology. They are not homophonous with their SG forms, and still carry one number cue each.

While a comparison of isolated nouns between WAS and Castilian is useful for analyzing lexical differences, more information may be gathered by comparing the structural differences in the gendered and numbered forms of the determiners in these dialects. When nouns are preceded by determiners, there is far less potential for ambiguity. The number marking differences created by the interplay of determiners and nouns can be exploited to create even more detailed structural ambiguity criteria, which can then be compared within and across dialects.

Determiners are a specific class of modifiers that give semantic information about nouns. In Spanish, determiners precede nouns, and give modifying information such as definiteness, quantity, possession, and location/deixis. The grammatical number and gender of a determiner depend on the number and gender of the noun it modifies, so the morphological shape of determiners varies. Table 2 is a model of how V-final and C-final nouns interact with the gender and number of determiners, to predict the quantity of

number cues available on a phrase when [s] is present and when [s] is absent. The determiners used in Table 2 are the definite articles [la] (FEM, SG), [las] (FEM, PL), [el] (MASC, SG), [los] (MASC, PL), all of which translate as "the." Note that there is a historical structural difference in the shape of most singular masculine determiners that is crucial to the current methodology.

Table 2

*The Effect of [s] Lenition on Spanish Plural Morphology*

| | Vowel-final Noun | | Consonant-final Noun | |
|---|---|---|---|---|
| | Feminine Determiner | Masculine Determiner | Feminine Determiner | Masculine Determiner |
| **Non-lenition**<br><br>Singular<br>Plural<br>Cues | [la amiɣa]<br>[la-s amiɣa-s]<br>2 | [el amiɣo]<br>[los amiɣo-s]<br>2 | [la muxeɾ]<br>[la-s muxeɾ-es]<br>2 | [el profesoɾ]<br>[los profesoɾ-es]<br>2 |
| **Lenition**<br><br>Singular<br>Plural<br>Cues | [la amiɣa]<br>[la amiɣa]<br>0 | [el amiɣo]<br>[lo amiɣo]<br>1 | [la muxeɾ]<br>[la muxeɾ-e]<br>1 | [el profesoɾ]<br>[lo profesoɾ-e]<br>2 |
| **Ambiguity potential** | Ambiguous | Not ambiguous | Not ambiguous | Not ambiguous |

While each phrase in Castilian has two cues to plurality (one on the determiner and one on noun), the reduced plural morphology of WAS nouns seen in Table 1 persists in the phrase environment shown in Table 2. In a lenition dialect, feminine determiners followed by feminine nouns may have no word- or phrase-level number cues to differentiate the singular from the plural form. The absence of number cues in this

specific environment is the key component when comparing dialects by their potential for ambiguity. In all the other DET+N environments, there is at least one differentiator, even when [s] is absent. For instance, the V-final noun with masculine determiners is differentiated by a historical structural difference in the masculine determiner morphology, such that [el] and [lo] in WAS are as easy to distinguish as [el] and [los] in Castilian. On the C-final side, the feminine [muxeɾ] has homophonous singular/plural determiners, but maintains a contrast with the remnant [e] seen earlier. Finally, the masculine, C-final [profesoɾ] has cues in both the determiner and in the remnant [e] of the noun.

Most other determiners follow a similar pattern as the definite articles, with structural differences in the masculine determiners creating less ambiguity potential.[6] Table 3 shows the Spanish determiners. Bolded determiners highlight those which are homophonous in singular and plural forms.

---

[6] Word-final [s] lenition additionally causes derived homophony between masculine determiners and neuter pronouns, but the differing syntactic placement of the pronouns prevents confusion.

Table 3

*Determiners in Castilian and WAS*

| Determiners | | SG CaS/WAS | CaS PL | WAS PL distinct | homophones | *Gloss* |
|---|---|---|---|---|---|---|
| Definite | *m* | el | los | lo | | *the* |
| Articles | *f* | **la** | las | | **la** | |
| Indefinite | *m* | un | unos | uno | | *a, an,* |
| Articles | *f* | **una** | unas | | **una** | *some* |
| Alt. Indef. | *m* | algún | algunos | alguno | | *some* |
| Articles | *f* | **alguna** | algunas | | **alguna** | *(alternate)* |
| Demonstrative | *m* | este | estos | esto | | *that,* |
| Adjectives | *f* | **esta** | estas | | **esta** | *those* |
| Demonstrative | *m* | ese | esos | eso | | *this, these* |
| Adjectives | *f* | **esa** | esas | | **esa** | |
| Demonstrative | *m* | aquel | aquellos | aquello | | *that/those* |
| Adjectives | *f* | **aquella** | aquellas | | **aquella** | *(distal)* |
| Possessive | *n* | **mi** | mis | | **mi** | *my,* |
| Adjectives | *n* | **tu** | tus | | **tu** | *your,* |
| | *n* | **su** | sus | | **su** | *his, her,* |
| | *m* | **nuestro** | nuestros | | **nuestro** | *your,* |
| | *f* | **nuestra** | nuestras | | **nuestra** | *their, our* |
| Quantifier | *m* | **mucho** | muchos | | **mucho** | *many* |
| (example) | *f* | **mucha** | muchas | | **mucha** | |
| Cardinals | | | | | | |
| (example) | *n* | **dos** | | | | *two* |

The fact that some nouns and determiners retain number-cues without [s], and some do not, offers a concrete way to compare the relevance of number marking across dialects, and forms the basis of the current study. In comparing these dialects, I am looking for evidence of large-scale compensation in WAS that implies that word-final [s] acts as a critical semantic contrast in Spanish, such that its loss requires some kind of active, synchronic compensation.

**1.4. The Questions to Be Addressed**

The literature review in Section 2 demonstrates that both phonetic compensation and the purposeful, strategic addition of [s] as a plural marker have been ruled out in WAS; however, there have not yet been studies addressing the possibilities of lexical or morpho-syntactic compensation for [s] lenition in this dialect. These must be addressed before a comprehensive discussion about the consequences of lenition can be attempted.

There are two potential means of lexical compensation that are investigated by this study, and four potential means of morpho-syntactic compensation which are investigated, across 60 DET+N phrase environments. The combined weight of the answers to these research questions form the justification for my broader claim about the relative semantic irrelevance of [s] in WAS.

The following questions are addressed in this study:

1. Is there a difference in the noun type-token ratios (TTRs) of WAS and Castilian? This calculation would give information about respective lexical complexity between these dialects. For instance, the potential number-ambiguity of some nouns in WAS might have led to an increase in the use of nouns that carry number marking inherently (words like *gente* "people" or *equipo* "team"), or some other manner of avoiding ambiguity through a more complex lexicon. A difference between the dialects in TTR could indicate lexical compensation in WAS.

2. Do WAS speakers use fewer V-final nouns (which are potentially ambiguous) than Castilian speakers? This would answer the question of whether there has been lexical restructuring in WAS, such as a tendency to default to

C-final synonyms, which are non-ambiguous, rather than V-final ones, which are ambiguous. In WAS, this might be observed as a frequency reduction of nouns which have zero number cues without [s]. For instance, across the many countries that speak Spanish, there are 17 different words for pen. A few include: *el boligrafo, la pluma,* and *el lápiz tinta*. Each of these examples would have a different quantity of recoverable number cues in WAS, and if grammatical number is highly relevant, there could be a preference in lenition dialects for the less ambiguous options.

3. Do WAS and Castilian speakers differ in the frequency of DET+N phrases when the phrases are compared by how many structural number-cues they carry when [s] is not present? Because DET+N phrases can vary between one, two, or zero remaining number-cues without [s], an avoidance in WAS of DET+N constructions that carry zero number-cues might indicate compensation for the loss of word-final [s].

4. Does WAS use more determiners than Castilian, compared to the use of bare nouns (nouns with no preceding determiner)? DET+N phrases are far more likely to carry number cues than bare nouns. An avoidance of bare nouns in WAS should be predicted if number marking is highly relevant.

5. Does WAS use more cardinal numbers or quantifiers than Castilian? These determiners carry number inherently, regardless of whether there are recoverable morphological number-cues or not. If number marking were unclear or a critical distinction, an increased use of these determiners would be expected.

To recall, I find WAS does not differ substantially from Castilian in ways that would point toward synchronic compensation for [s] lenition, but there may be evidence in the feminine indefinite articles suggesting diachronic changes in the relevance of [s] at a previous stage of the language. The evidence in this corpus comparison of WAS and Castilian points to a low semantic relevance of word-final [s] in Modern Spanish.

The rest of this thesis includes a literature review that covers Spanish [s] lenition studies, as well phonological studies relevant to [s] lenition (Section 2). Data and methodology are discussed in Sections 3 and 4. Section 5 has the results of comparing nominal number cues in WAS and Castilian, and Sections 6 and 7 are the Discussion and the Conclusion.

## 2. Literature Review

This literature review serves two purposes. The first is to provide evidence from quantitative cross-linguistic phonological studies that show that if word-final [s] were still a critical contrast in Spanish, it would necessarily persist or trigger compensation. The second is to present evidence from the literature specific to WAS [s] lenition, demonstrating that word-final [s] is no longer a plural marker in WAS, and that it has not triggered compensation in any of the previously tested environments.

### 2.1. Relevant Contrasts Tend to Persist

To answer the question of whether [s] can undergo lenition due to a lack of semantic relevance, it must first be established that relevant contrasts tend to either persist or be compensated for, and that it would be unusual for this sound to undergo lenition without compensation, if it is shown to be relevant.

As mentioned in the introduction, Gurevich (2004) investigates lenitions in 153 languages and finds that 92% of lenitions avoid phonological neutralization. Gurevich makes a distinction between *phonetic* neutralization and *phonological* neutralization. Any sounds that neutralize, such that two sounds lose their contrastive status with respect to each other, can be said to have undergone phonetic neutralization. This kind of neutralization may or may not affect semantic clarity. If the neutralization creates homophony, and results in the loss of communicative function, then according to Gurevich's criteria, it would be a phonological neutralization.

Silverman (2010) provides data that support Gurevich 2004 by investigating six different neutralizations in Korean, and quantifying the number of minimal pairs they

induce. Silverman finds that rampant neutralization is free to proceed in Korean because the neutralizations do not tend to create homophones, and therefore do not create semantic constraints. Like those in Gurevich 2004, Silverman's results suggest that neutralizations that result in semantic ambiguity are unlikely to proceed.

These two studies offer a lot of quantitative data suggesting that the likelihood that a sound will undergo lenition is not independent of that sound's functional load. Martinet (1952) originally describes functional load primarily as a measure of the quantity of minimal pairs for which a sound is responsible. He laments:

> It is clear that the functional yield of an opposition can only be evaluated with any degree of accuracy if we deal with the linguistic stages for which fairly exhaustive word lists are available. This circumstance makes it practically impossible to check the validity of the functional assumption in the case of prehistoric sound shifts. (Martinet, 1952, p. 9)

The widespread availability of online data has changed this circumstance somewhat. This is particularly true in Spanish, which not only has a long history of linguistic inquiry and written records, but can also be traced backwards and forwards in dozens of dialects from the splits of Peninsular dialects, to the linguistic effects of colonialism, to modern-day dialects.

The ability to analyze larger data sets has allowed contemporary researchers to test Martinet's criteria. Wedel, Kaplan, and Jackson (2013) utilize corpus linguistics in the development of modern criteria for functional load. This eight-language study identifies trends that sounds with high functional load tend to share, and confirms that sounds that meet the criteria for high functional load are less likely to merge (or neutralize).

According to the criteria of this study, phonemes that are responsible for a large number of minimal pairs are less likely to have merged over the course of time.

At first glance, it seems like word-final [s] should be a candidate for high functional load since it is responsible for distinguishing so many singular nouns from plural nouns. However, Wedel, Kaplan, and Jackson also demonstrate that high-frequency sounds tend to have a low functional load. It is well established that [s] is the highest frequency consonant by far in Castilian (Sandoval, Toledano, De La Torre, Garrote, & Guirao, 2008). This is a data-based indication that [s] may not be as important as previously believed after all.

If we apply the findings of Gurevich (2004), Silverman (2010), and Wedel, Kaplan, and Jackson (2013) to Spanish [s] lenition, we see the following progression: [s] undergoes phonetic neutralization toward zero, deriving singular/plural homophony. Without further information, we would expect the semantic contrast provided by word-final [s] to be compensated for in some manner. If no compensation can be found for this neutralization, we must conclude that word-final [s] does not have a high functional load in Modern Spanish, and that it should be considered a phonetic neutralization, not a phonological one.

## 2.2. Word-Final [s] Is not a Plural Marker in WAS

The combined results of the following studies strongly suggest that word-final [s] does not function as a plural marker in WAS. There is simply no evidence that WAS speakers use word-final [s] (or any other acoustic variants) contrastively, or that there is any need for [s] as a plural marker in WAS.

**2.2.1. Listeners cannot distinguish plural nouns from singulars in WAS.** The data

analyzed in O'Neill 2005 are provided by WAS speakers from Málaga and Seville. The

intention of the study is to uncover possible phonetic compensation for morphological [s]

lenition. O'Neill discusses the commonly asserted idea that Eastern Andalusian (EAS)

speakers use vowel harmony to distinguish plurals (see Hernandez-Campoy and Trudgill

2002 for a refutation), and seeks to document an acoustic compensatory strategy in WAS,

perhaps with contrastive aspiration on plurals.

O'Neill recorded six speakers producing singular and plural words. He then edited the

sound files to isolate the nouns from the determiners, and the pronouns from the verbs.

He mixed them up and played the isolates back to the same set of speakers. Each subject

listened to a combined list of 176 words and attempted to label words as either

singular/plural or 1st person/2nd person. They were informed that there were duplicates

in the data. The percentages of correct answers for each speaker were: 40%, 46%, 56%,

40%, 49%, and 46%. In sum, they were not able to distinguish isolated plurals or isolated

verbal inflections by ear. They performed worse than chance.

Furthermore, it was found that when the informants listened to their own voices

producing the isolated singulars and plurals, the results were the same; they could not

differentiate the isolated plurals from singulars. The WAS listeners were not able to use

any sub-segmental [s] remnants to identify the isolated plurals.

While the results of O'Neill 2005 point toward no phonetic compensation for nouns

in isolation, that study does not address the possibility that there may be durational cues

that persist on vowels (after [s] lenition), which might be perceived between determiners and nouns or nouns and the next word. That possibility is explored in the next study.

**2.2.2. Word-final duration is not a plural cue.** Carlson (2012) also performs plural identification tests, but this time includes ambiguous nouns in ambiguous carrier phrases. The new information provided by looking at nouns within phrases, is that listeners are not able to use word-final vowel duration as a cue to plurality.

Six WAS speakers are recorded producing sentences and words, and 25 Andalusian listeners attempt to disambiguate the words only by phonetics. Carlson uses consultants from geographic regions that are traditionally said to have both WAS speakers and EAS speakers, but acoustic analysis of her consultants' speech did not find any vowel lowering, which in older speakers is said to be a hallmark of the EAS accent. There are no differences in the results based on region of origin.[7]

The consultants are tested on words that are isolated from other cues, and on words in numerically ambiguous carrier phrases. As in O'Neill 2005, it is found that consultants cannot identify plural nouns from singulars. In the various identification tests with singular and plural nouns and phrases, the ability to correctly recover morphological information ranged from 54.1% to 57.9% for ambiguous nouns in ambiguous carrier phrases. This marks a higher success rate than that seen in O'Neill 2005, but it is still not high enough to demonstrate phonetic disambiguation. The fact that WAS consultants do not recognize number-ambiguous plurals when they are placed in ambiguous carrier phrases suggests that duration is not a contrastive cue in the word-final environment.

---

[7] Regions of origin established via personal correspondence with Carlson.

O'Neill (2005) and Carlson (2012) make a compelling case so far against phonetic compensation. The next study is used to cross-check the previous results from a different perspective.

**2.2.3. Children must learn to write word-final < s >.** The possibility of phonetic compensation for [s] lenition in WAS appears even less likely in the work of Defior et al. (2008), which is an educational study from Spain which focuses on how native-WAS-speaking children use both phonology and morphology as they are learning to spell in Spanish. The aim of the study is to test the potential role of morphology in Spanish spelling. Of relevance to the current study, word-final <s> is chosen as the test grapheme because the authors agree that the WAS school children do not have a corresponding phoneme (in the word-final environment) to represent this grapheme, and that no phonetic compensation for the sound has been demonstrated in this dialect. They note that children in Andalusia must be taught to add an orthographic <s> to the end of written plurals, (just as English-speaking children must be taught to add a silent <e> to some words). "The participants' tendency to write down final <s> was rather modest, no more than 40% in 1st grade and about 65% in 3rd grade. This shows that the phoneme /s/ was indeed absent from the input as well as from the internal lexicon of the participants." (Defior et al., 2008, p. 211)

In the study, children listened to sentences recorded by consultants who do not have word-final [s] in their phonetic inventory. To absolutely rule out phonetic cues, the recordings are digitally checked and cleaned so that the small percentage of aspirations or other remnant sounds are removed. Recall that there is already evidence in this dialect

that these sounds are not perceived as plural markers by WAS speakers. The sentences have contextual and grammatical cues in them that make them obviously singular or plural, for example: *La(s) manzana(s) son buena(s) para la salud… manzana(s)* ("Apples are good for health… apples"). The verb in this sentence is plural, and in this case provides enough information to pluralize the rest of the sentence. There is no indication from the authors that there is any confusion over the semantic content of the simple sentences. The third-grade students show consistent increases in orthographic plural production in all categories over the first graders.

At the age of nine years old, the WAS third graders are unlikely to be improving in their ability to understand the concept of plurality, but rather they are simply learning to spell. Marrero and Aguierre (2003) show that Spanish-speaking children in [s] retaining dialects acquire morphologic plural marking by about two years old, and Miller (2007) finds that children in variable lenition dialects may not acquire word-final [s] until the age of 4 or 5 years old. In a dialect in which [s] in almost never produced, it should not be surprising that orthographic <s> would need to be actively taught. The key point here is not about orthography however; the useful information provided by this study is that the children are not only learning to spell, but also (incidentally) demonstrating that they gather information about plurality by way of context. Each time they correctly add an orthographic <s> to a V-final plural noun, it is evidence that they are using something other than nominal plural morphology to make that decision.

To see this in action with adults, one need only look at the orthographic transcripts of this study. Adult Spanish transcribers identify plural nouns in the WAS corpus by writing

an orthographic <s> at a ratio comparable to plurals that were identified in the Castilian corpus. Most of those written tokens are not produced in the spoken WAS corpus, and yet in 36 conversations, across two dialects, native speaker transcribers in WAS and Castilian identified virtually the same ratio of singular/plural noun use. WAS listeners do effectively de-code grammatical number in Spanish.

**2.2.4. Number is overtly signaled 94% of the time in WAS.** Ranson 1993 is an unprecedented and exhaustive semantic study in which the many different ways that plurality is redundantly cued in Spanish are documented. Ranson concludes that WAS speakers do not experience function-negative consequences due to [s] lenition because when contextual and structural plural cues are all considered (such as verbal inflection or determiner cues), only 6% of nouns are marked by neither morphology nor context (Table 4). Of that 6%, none results in semantic ambiguity as they can be disambiguated via pragmatic inference.[8]

---

[8] This has been my experience in spot-checking the current corpora as well. When context is taken into account, I was unable to find nouns that are truly semantically ambiguous in either corpus.

Table 4

*Semantic and Morphological Number Marking of WAS Nouns*

| The noun is:<br>(1627 nouns) | Marked by morphology<br>(on noun or verb phrase) | Not marked by<br>morphology | Total |
|---|---|---|---|
| Number Irrelevant | 24% | 16% | (40%) |
| Marked by context | 32% | 13% | (45%) |
| Not marked by context | 9% | **6%** | (15%) |
| **Total** | 65% | 35% | (100%) |

*\*Note*. From "The interaction of linguistic and contextual number markers in Andalusian Spanish," by D.L. Ranson, 1993, *Hispania, 76*(4), p. 930. Copyright 1993 by Diana Ranson. Adapted with permission.

It is further found that the Western Andalusian adults in the study do produce word-final [s] 1.5% of the time, and may produce an allophonic [h] or aspiration up to 4.5% of the time. However, none of these sounds is produced in any patterned way, and there is no evidence that they are used in a compensatory manner, or that speakers are making any real-time decisions about how to distinguish plurals for their listeners. Ranson 1993 documents a flexible and redundant WAS plural system in which grammatical number is signaled variously and simultaneously across the nominal morphology, in verbal inflection, and especially, contextually throughout the discourse.

As previously discussed, O'Neill 2005 and Carlson 2012 effectively rule out phonetic compensation in WAS. The results of Defior et al. 2008 support those studies and provide evidence that WAS children do not require [s] to de-code plurality, such that the larger context and cues in a sentence are enough for comprehension. Ranson 1993 shows that there is no pattern to the occasional use of word-final [s] in WAS, and that it is wholly unnecessary to speakers' and listeners' purposes.

## 3. The Data

The data for this study are from 18 recorded interviews in the publicly available PRESEEA-MA corpus (Málaga Urban Vernacular, 2007; WAS dialect) and 18 from the publicly available PRESEEA corpus (PRESEEA, 2014; Madrid, Castilian). The recordings were made between 1990 and 1994. They were transcribed by native-speaker linguists between 2000 and 2008. The PRESEEA interviews were chosen to match the already balanced Málaga corpus (Table 5), which has equal representation of gender, three age groups (20-34, 35-54, and 55+), and three education levels (primary, secondary, university).

The interviews range from 30 minutes to an hour. In pursuit of balanced demographics, it was necessary to sacrifice parity in the sizes of the dialect samples. The Madrid corpus turned out 15% larger than the Málaga corpus, and so ratios, percentages, and Chi-square analyses are used to compare type and token frequencies throughout this thesis.

Table 5

*Consultant Demographics in the Madrid and Málaga Corpora*

| Castilian Speakers (Madrid) | | | | WAS Speakers (Málaga) | | | |
|---|---|---|---|---|---|---|---|
| Speaker | Edu Lvl. | Gender | Age | Speaker | Edu Lvl. | Gender | Age |
| CS 01 | lower | M | 20 | WAS 01 | lower | M | 24 |
| CS 02 | lower | F | 21 | WAS 02 | lower | F | 23 |
| CS 03 | lower | M | 42 | WAS 03 | lower | M | 41 |
| CS 04 | lower | F | 42 | WAS 04 | lower | F | 45 |
| CS 05 | lower | M | 71 | WAS 05 | lower | M | 64 |
| CS 06 | lower | F | 75 | WAS 06 | lower | F | 80 |
| CS 07 | medium | M | 33 | WAS 07 | medium | M | 32 |
| CS 08 | medium | F | 21 | WAS 08 | medium | F | 30 |
| CS 09 | medium | M | 35 | WAS 09 | medium | M | 52 |
| CS 10 | medium | F | 52 | WAS 10 | medium | F | 43 |
| CS 11 | medium | M | 65 | WAS 11 | medium | M | 61 |
| CS 12 | medium | F | 65 | WAS 12 | medium | F | 61 |
| CS 13 | high | M | 29 | WAS 13 | high | M | 28 |
| CS 14 | high | F | 22 | WAS 14 | high | F | 28 |
| CS 15 | high | M | 39 | WAS 15 | high | M | 50 |
| CS 16 | high | F | 51 | WAS 16 | high | F | 39 |
| CS 17 | high | M | 75 | WAS 17 | high | M | 60 |
| CS 18 | high | F | 75 | WAS 18 | high | F | 56 |

*Note.* Edu Lvl = Education Level

# 4. Methodology

This section includes the steps that were taken to prepare the data for analysis, and the methodology used to organize and identify the patterns in the data.

## 4.1. Data Preparation

For this study I performed part-of-speech (POS) tagging on each corpus using the tool Tree Tagger (Schmid, 1995). I cross-checked the tags syntactically using MS Excel filtering, and corrected tags where necessary. In quality checks of the tagging, the determiner tagging was found to have no errors, while the noun tagging was found to be 98% accurate after corrections, with most errors represented by modifier nouns being marked as argument nouns, which has no identifiable effect on the final results. Common nouns were extracted from the transcripts along with their attendant determiners. Determiner repeats, stutters, and repairs were excluded, such that only the last determiner uttered before a noun is collected.

Every noun in the corpora has identifying information attached that makes it possible to filter, sort, or view the noun in context, for the purpose of verifying syntactic or semantic role. Each noun is tagged by whether its singular form is V-final or C-final. Nouns are also tagged for the gender of any determiner that may precede them, as well as number, if it is identifiable without /s/.

This study is about the spontaneous use of language; therefore, all the interviewer questions and utterances were removed prior to the frequency counts, and only consultant speech was analyzed. The transcribers of the Madrid data orthographically transcribed every utterance, including filler noises like <ah> and <eeeee>. The Málaga corpus does

not include these sounds. These sounds were excluded, and the final word count was calculated as 126,515 words for the Madrid corpus, and the 108,695 words for the Málaga corpus.

The transcripts are orthographic, and they provide an accurate record of the lexical and morpho-syntactic choices of the WAS and Castilian consultants, and it is these choices that form the basis of this dialect comparison, rather than phonetic distinctions.

**4.1.1. Nouns.** In these pages, "noun" refers to a common lexical noun acting as an argument of the predicate, and not in a modifying capacity. Therefore, in the sentence, "*Juan es jugador de tenis"* ("Juan is a tennis player"), *jugador* would be analyzed as a C-final, singular, masculine common noun with no preceding determiner, but *tenis* would not because it is acting in the role of a modifier, rather than an argument. Modifying nouns are subject to the same kind of singular/plural homophony, and may or may not be preceded by determiners, but these are usually either number irrelevant or disambiguated by verbal inflection or context. An adjective or a cardinal number can also serve in a noun capacity (*el grande* "the big one" or *los tres "*the three of them"), but for the purposes of this inquiry, they are also excluded.

Proper nouns were excluded from this investigation. In cases in which a distinction needed to be made between a phrase like (a) *el mercado Alvarez* "Alvarez Market" and (b) *el mercado de los Alvarez "*the Alvarez's market', example (a) was excluded as a proper noun. In (b), the proper noun is only modifying *el mercado*, therefore *el mercado* is retained, and only *los Alvarez* was excluded.

In an initial corpora comparison using frequency profiling (Rayson & Garside, 2000), I found three high frequency nouns that stood out in terms of both their frequency and the inequity of their use. These are *barrio* "neighborhood," *cosa* "thing," and *hombre* "man." After further investigation, all three were excluded from the analysis, in both corpora, in singular and plural forms. They are unevenly overused in ways that do not contribute nominal meaning. *Barrio* is used significantly more by Castilian speakers (Table 5), who use it multiple times in discussing where they lived as children. The WAS speakers predominantly use *calle* ("street") instead, when discussing locations, and most instances of *calle* were already excluded due to being proper nouns. For instance, *la calle 21* ("21st Street") is considered a proper noun, and is excluded. Hence the combination of overuse, and the inequity of use, resulted in the decision to exclude *barrio* as well. *Cosa* is significantly overused in WAS in idiomatic expressions, rather than as a true nominal. *Hombre* is used as a filler word in Castilian, but not in WAS.

Table 6

*Token Frequencies of Excluded Nouns*

| Noun | Castilian | WAS |
|---|---|---|
| *hombre* | 274 | 103 |
| *cosa* | 242 | 430 |
| *barrio* | 210 | 55 |
| *verdad* | 114 | 144 |

There is one additional exclusion to be discussed in Table 6. The noun *verdad* ("truth") is used as a discourse marker in both corpora, to mean "Right?" It is also used to signal agreement in the expression, "*Es la verdad*" ("It is true") in WAS, or "*Es verdad*"

*("It is true")* in Castilian. The omission of *la* in Castilian does not affect semantic clarity in any way. The imbalance of determiner usage (Table 7), and the fact that *verdad* is not functioning in the role of a lexical noun, requires exclusion. *Verdad* is the only noun in the category of "preceded by a feminine definite article" to have a 47-token difference in WAS. The other nouns in that category differed in frequency by 1 to 10 tokens.

Table 7

*Token Frequencies of the Noun Verdad*

| Preceding *verdad* | Castilian | WAS |
|---|---|---|
| Determiner (FEM DEF ART: "la/las") | 53 | 100 |
| No Determiner | 61 | 44 |
| Total | 114 | 144 |

The exclusion of these four words does not make the two corpora incompatible in terms of complexity. Table 8 compares three configurations of nouns in the corpora by type and token ratios. In (a), the ratio of type and token, for all nouns, including modifiers, and all four over-use words, is calculated for both dialects. Castilian has a type-token ratio of 0.178, and the WAS ratio is 0.179. By this comparison, the dialects are of similar complexity. The same type and token counts are compared between dialects using Chi-square testing, with the results $\chi^2(1)=0.004$, p=0.952. There is no statistically significant difference between the dialects by this measure.

Table 8

*Nouns by Type, Token, and Ratio*

| Comparisons According to Various Criteria | Castilian | WAS |
|---|---|---|
| a.  All nouns, including overused nouns, modifiers | | |
| Type | 2919 | 2553 |
| Token | 16364 | 14287 |
| Ratio | 0.178 | 0.179 |
| b.  All Nouns, -barrio, -cosa, -hombre, -verdad | | |
| Type | 2915 | 2550 |
| Token | 15512 | 13547 |
| Ratio | 0.188 | 0.188 |
| c.  Nouns, -modifiers; -barrio, -cosa, -hombre, -verdad | | |
| Type | 2759 | 2418 |
| Token | 14485 | 12623 |
| Ratio | 0.191 | 0.192 |

In (b), the same comparisons are done, but this time excluding the over-use nouns but including the modifiers. The difference between type-token ratios is still negligible, and the chi-square results are also not significant: $\chi^2(1)=0.003$, p=0.959.

In deciding which noun configuration to use for the deeper analysis, option (c) was chosen from Table 8. This option includes lexical nouns, but excludes modifiers and the over-use words. The type-token ratio for Castilian is 0.191 and for WAS is 0.192. These are comparable, and the chi-square analysis is $\chi^2(1)=0.035$, p=0.852, which is not significant. The exclusion of *barrio, hombre, cosa, and verdad*, as well as the modifier nouns, are made from an abundance of caution, and the desire to isolate and control variables as much as possible.

**4.1.2. Determiners.** Determiners are extracted from the corpora in conjunction with the noun they modify. Determiners are found in three positions, relative to a noun. "Noun -1" is the most common position for determiners, as in *el amigo*, but "Noun -2" is possible when there is an adjective modifying the noun. In that case, the most common sequence is "Determiner + Adjective + Noun." While the most common location for an adjective is in the post-noun position, there are examples in the data with an adjective in the Noun -1 slot. In all cases, nouns trigger gender and number marking on the other elements of the phrase, and including adjectives is unnecessary for the current investigation. For the purposes of this study, the phrase *el buen hombre* (the good man) would be processed as *el hombre*. Note that the truncated masculine adjective *buen* does cue number, but it is completely redundant to the other two cues already in the phrase.

While less frequent, it is also possible to have determiners in both the Noun -2 and Noun-1 positions. For example, _las dos hermanas_ ("the two sisters"), or _todos los hermanos_ ("all the brothers"). Because double determiners most often have a cardinal or quantifier in the phrase, they are rarely ambiguous for number. These are assessed separately from the single determiners in order to rule out WAS using double determiners as a disambiguating strategy.

**4.2. The Study: Comparing WAS and Castilian**

All 27,366 nouns, from both dialects, are identified by the number of recoverable plural cues (one, two, or zero) that a listener could be expected to recover if [s] is absent. These are the same cues presented in Section 1.3, Table 1. In the lexical tests, nouns are assessed in isolation, without their determiners. The type and token frequencies of the

isolated nouns provide information about the lexical choices of speakers, irrespective of the larger syntactic environments in which they are used. In the morpho-syntactic tests, nouns and determiners are analyzed together, as a contextual unit, exactly as they were extracted from the corpora. Each DET+N phrase is categorized by recoverable number cues, as shown in Section 1.3, Table 2.

After the isolated nouns are categorized and coded according to the structural ambiguity criteria of one or zero recoverable cues without [s], they are compared by type, token, and ratio across dialects. Next, the nouns are compared by the determiner phrase environment in which they were spoken. The identifying criteria in Table 9 are used on all the DET+N phrases in each corpus (including those with Ø determiner).

Table 9

*Number Cues on DET+N Phrases Without [s]*

| The singular form of the noun is: | With preceding: | Number Cues without [s] |
|---|---|---|
| Vowel-final | No Determiner | 0 |
| Vowel-final | Feminine Determiner | 0 |
| Vowel-final | Masculine Determiner | 1 |
| | | |
| Consonant-final | No Determiner | 1 |
| Consonant-final | Feminine Determiner | 1 |
| Consonant-final | Masculine Determiner | 2 |

The comparison works because we know that Castilian speakers reliably produce [s], while WAS speakers rarely do, and never in patterned ways (Ranson 1993). If there is lexical or morpho-syntactic compensation, it should present itself as a difference in the frequency between the dialects, of structurally ambiguous nouns or determiner phrases.

## 5. Results

The results presented here reference the five questions presented in Section 1.4. Questions 1 and 2 are lexical comparisons of WAS and Castilian. Questions 3-5 are morpho-syntactic questions. There is one statistically significant result in the data. It is found in the feminine indefinite articles (una, unas).

### 5.1. No Difference in Lexical Complexity

Question 1: Is there a difference in the noun type-to-token ratios of WAS and Castilian? In Table 8 of section 4.1.1, it was shown that the type and token ratios for nouns in these dialects are comparable, suggesting the dialects do not differ in lexical complexity. Indeed, the chi-square test of independence suggests that the ratios are not significantly different between the two: $\chi^2(1)=0.005$, p=0.940. Speakers of WAS do not appear to have a larger inventory of lexical nouns than speakers of Castilian.

### 5.2. No Evidence of Lexical Restructuring

Question 2: Do WAS speakers use fewer V-final nouns than Castilian speakers, or more C-final nouns? Castilian and WAS do not differ in the type or token frequencies of structurally ambiguous nouns (V-final) versus unambiguous ones (C-final), despite 79.42% of nouns in WAS carrying no number morphology if [s] is absent. Chi-square analysis (Table 10) demonstrates that WAS and Castilian noun frequencies are comparable when compared by V-final and C-final nouns, by type: $\chi^2(1)=0.877$, p=0.349.

Table 10

*Type Frequencies of Vowel-Final and Consonant-Final Nouns*

| Nouns by Type | CaS | WAS |
|---|---|---|
| V-final | 2028 | 1805 |
| C-final | 731 | 613 |
| Total | 2759 | 2418 |

*Note.* CaS = Castilian

When C-final and V-final nouns are compared by token frequency, the difference is also not significant: $\chi^2(1)=0.275$, p=0.599 (Table 11). There does not appear to be any large-scale lexical adjustment away from nouns that become homophonous in singular/plural (feminine determiners followed by vowel final nouns), or that differ in number of cues without word final [s]. These dialects appear to be highly comparable in the structural composition of their lexicons, despite regional lexical differences recorded in the data.

Table 11

*Token Frequencies of Vowel-Final and Consonant-Final Nouns*

| Nouns by Token | CaS | WAS |
|---|---|---|
| C-final | 2800 | 2472 |
| V-final | 11685 | 10151 |
| Total | 14485 | 12623 |

This does not mean that these dialects use the exact same words. It means that they use the same kinds of words. WAS speakers show no preference for nouns that are less structurally ambiguous. Table 12 provides an example in which speakers of WAS and speakers of Castilian have different preferences for the words they use to talk about their

children. While both dialects use all four terms, WAS speakers strongly prefer forms of

*niño* to talk about children, but Castilian speakers show more use of *hijo* and *chico*.[9]

Table 12

*Nouns Used for Children in Castilian and WAS*

| Some nouns used to talk about children | Castilian | WAS |
|---|---|---|
| *niño/s* | 45 | 137 |
| *niña/s* | 23 | 42 |
| *chico/s* | 46 | 14 |
| *chica/s* | 38 | 12 |
| *hijo/s* | 74 | 26 |
| *hija/s* | 79 | 16 |
| *chaval* | 13 | 1 |
| *chavale/s* | 2 | 0 |

The lexical diversity demonstrated here is clearly due to something beyond ambiguity

because in the majority of these forms there are no more number cues than any of the

others. Note that the only option that ends in a consonant, and is not therefore

numerically ambiguous (*chaval*), is used almost exclusively by Castilian speakers, who

pronounce [s]. The lexical choices may hinge on many different factors, but avoiding

word-level number ambiguity is not one of them.

## 5.3. A Frequency Difference in DET+N Phrases by Number Cues

Question 3: Do WAS and Castilian speakers differ in the frequency of DET+N

phrases when the phrases are compared by how many structural number-cues remain

when [s] is not present? There are no significant frequency differences found between

---

[9] The masculine form (o-final) is the default for mixed groups, and so has a higher
frequency.

WAS and Castilian in 98.7% of DET+N phrases. The statistically significant difference is found in phrases of the construction: FEM INDEF ART + NOUN (Table 16).

The process for isolating the significant result begins with organizing the DET+N phrases by the categories established in Table 9 (Section 4.2). Table 13 is based on the key for calculating number cues on DET+N phrases, without [s], as shown in Table 9. The data are arranged in all the possible DET+N configurations, and each configuration has a predictable number of plural cues without [s]. The token totals include all the nouns, as preceded by the nine determiners in Table 3, plus the "No Determiner" category. There are 10 determiner options and six phrase environments, for a total of 60 individual DET+N configurations.

Table 13

*Distribution of DET+N Phrase Tokens in Castilian and WAS*

| Determiner Gender | Noun Ending | Cues w/o [s] | CaS | WAS |
|---|---|---|---|---|
| No Determiner/Neuter | V-final | 0 | 4773 | 4013 |
| Feminine Determiner | V-final | 0 | 3300 | 2772 |
| | | | | |
| Masculine Determiner | V-final | 1 | 3551 | 3322 |
| No Determiner/Neuter | C-final | 1 | 1085 | 874 |
| Feminine Determiner | C-final | 1 | 911 | 887 |
| | | | | |
| Masculine Determiner | C-final | 2 | 865 | 755 |
| Total | | | 14485 | 12623 |

Chi-square analysis of all six determiner environments in Table 13 has a result of $\chi^2(5)=22.006$, p=0005, which is significant. Narrowing the DET+N totals to compare

them by the number of cues (Table 14) yields $\chi^2(2)=11.533$, p=0.003, which is also significant as the following table shows.

Table 14

*Chi-Square of Total DET+N Tokens in Castilian and WAS*

| DET+N Phrases with: | Chi-square |
|---|---|
| Zero Cues | |
| 1 cue | $\chi^2(2)=11.533$, p=0.003 |
| 2 cues | |

*Note.* Comparison is by number-cues and dialect

To find which variable/s is/are causing the significant result, the determiner categories are compared by V-final and C-final nouns, by DET gender (Table 15). There is a significant result in the Feminine Determiner category, $\chi^2(1)=7.553$, p=0.006, but no significant results for the No Determiner/Neuter category, $\chi^2(1)=0.726$, p=0.394, nor the Masculine Determiner category, $\chi^2(1)=1.570$, p=0.210.

Table 15

*Chi-Square of DET+N Tokens by Gender*

| Determiner | Noun | Cues w/o [s] | Chi-Square |
|---|---|---|---|
| No Determiner/Neuter | V-final | 0 | |
| No Determiner/Neuter | C-final | 1 | $\chi^2(1)=0.726$, p=0.394 |
| | | | |
| Feminine Determiner | V-final | 0 | |
| Feminine Determiner | C-final | 1 | $\chi^2(1)=7.553$, p=0.006 |
| | | | |
| Masculine Determiner | V-final | 1 | |
| Masculine Determiner | C-final | 2 | $\chi^2(1)=1.570$, p=0.210 |

With the significant results isolated to feminine determiners, the next step is to isolate which feminine determiners have statistically significant differences when followed by a V-final noun, compared to a C-final noun.

I find that the significant results are in the FEM INDEF ARTs (*una/unas*): $\chi2(1)=8.619$, p=0.003 (Table 16). C-final nouns appear less often than expected (207 vs. 232.88) in Castilian while they appear more often than expected (195 vs. 169.12) in WAS. On the other hand, V-final nouns appear more often than expected (918 vs. 892.12) in Castilian while they appear less often than expected (622 vs. 647.88) in WAS.

Table 16

*Chi-Square of all FEM DET Tokens*

| Determiners | DET Gender | Noun (SG) | CAS | WAS | Chi-Square C-final v. V-final |
|---|---|---|---|---|---|
| Quantifiers | FEM | C-final | 173 | 190 | |
| | FEM | V-final | 246 | 213 | $\chi^2(1)=2.858$, p=0.091 |
| | | total | 419 | 403 | |
| Possessive Adjectives | FEM | C-final | 6 | 6 | |
| | FEM | V-final | 5 | 8 | $\chi^2(1)=0.337$, p=0.561 |
| | | total | 11 | 14 | |
| Dem. Adj. (3 sets) | FEM | C-final | 42 | 33 | |
| | FEM | V-final | 145 | 150 | $\chi^2(1)=1.122$, p=0.289 |
| | | total | 187 | 183 | |
| Indefinite Articles | FEM | C-final | 207 | 195 | |
| | FEM | V-final | 918 | 622 | $\chi^2(1)=8.619$, p=0.003 |
| | | total | 1125 | 817 | |
| Definite Articles | FEM | C-final | 483 | 463 | |
| | FEM | V-final | 1986 | 1779 | $\chi^2(1)=0.867$, p=0.351 |
| | | total | 2522 | 2342 | |

*Note.* Compared by C-final, V-final nouns and dialect. "Dem. Adj." = Demonstrative Adjetives, and includes all the *ese, este, aquel* forms

In the FEM INDEF ART + N phrases, the WAS corpus has 12 fewer phrase tokens than the Castilian corpus. Recall that the Castilian corpus is 15% larger than the WAS corpus, so this difference would be expected to be larger. In Table 17, a frequency comparison of FEM INDEF ART + C-final N phrases shows that the largest difference between WAS and Castilian is in the phrase *una habitación* ("bedroom"), which is used 14 times in Castilian and only twice in WAS. None of the phrases in Table 17 is ambiguous for number, and the most likely explanation for this is that WAS speakers use a different term for "bedroom." The second biggest frequency difference is in *una vez* ("one time" or "once"). This is notable for the number of occurrences, and the fact that most speakers are probably using this as part of an idiomatic expression that is not nominal. In future study, it should probably be excluded. The small number of these phrases, and the relatively small number of differences between the dialects leaves some room to question how functionally significant this result is. This will be discussed in Section 6.0.

Table 17

*Most Frequent C-Final Nouns Following a FEM INDEF ART*

| C-final noun preceded by FEM IND ART (una/s) | | Phrase tokens | | |
|---|---|---|---|---|
| | | Castilian | WAS | Difference |
| *habitación* | "bedroom" | 14 | 2 | 12 |
| *vez* | "time, occurance" | 16 | 26 | 10 |
| *ciudad* | "city" | 8 | 15 | 7 |
| *exploración* | "exploration" | 0 | 4 | 4 |
| *cantidad* | "quantity" | 4 | 7 | 3 |
| *facultad* | "faculty" | 0 | 3 | 3 |
| *instalaciones* | "installations" | 0 | 3 | 3 |
| *lesión* | "lesion" | 0 | 3 | 3 |
| *satisfacción* | "satisfaction" | 0 | 3 | 3 |
| *valoraciones* | "assessments" | 0 | 3 | 3 |
| *excursión* | "excursions" | 1 | 3 | 2 |

*Note*. FEM INDEF ART = una or unas

In the FEM INDEF ART + V-final N phrases, the WAS corpus has 296 fewer phrase tokens than the Castilian corpus. There are no other categories with a difference this large. A frequency comparison of the ten most frequent phrases in this category, shows that a full 79 tokens of that difference hinges on only two words: *persona* and *casa* (Table 18). The rest of the frequency differences are distributed throughout the nouns in this category. This will be addressed further in the discussion.

Table 18

*Most Frequent V-Final Nouns Following a FEM INDEF ART*

| V-final noun preceded by FEM IND ART (una/s) | | Phrase Tokens | | |
|---|---|---|---|---|
| | | Castilian | WAS | Difference |
| *persona* | "person" | 62 | 22 | 40 |
| *casa* | "house" | 59 | 20 | 39 |
| *zona* | "zone" | 22 | 7 | 15 |
| *chica* | "girl" | 18 | 4 | 14 |
| *copa* | "drink" | 13 | 2 | 11 |
| *especie* | "type" | 15 | 6 | 9 |
| *moto* | "motorcycle" | 9 | 1 | 8 |
| *vuelta* | "turn, return" | 11 | 4 | 7 |
| *manera* | "way, manner" | 10 | 3 | 7 |
| *plaza* | "plaza" | 8 | 1 | 7 |

## 5.4. Comparable Frequency of Bare Nouns

Question 4: Does WAS use more determiners than Castilian, compared to the use of bare nouns? WAS does not have a statistically significant frequency difference in determiner use compared to Castilian, despite a 38% increase in number cues when determiners precede nouns. In WAS, 77.1% of nouns are preceded by one of the following determiners: definite or indefinite articles, demonstrative adjectives, possessive adjectives, cardinals, and quantifiers. In the Castilian corpus, 77.5% of nouns are preceded by one of these determiners. Chi-square analysis finds no significant difference between the dialects in this regard $\chi^2(1)=1.392$, p=0.238 (Table 19).

Table 19

*Token Frequencies of Nouns Without Determiners*

| All Lexical | Tokens | | Percentage | | Chi-Square |
|---|---|---|---|---|---|
| Nouns | Castilian | WAS | Castilian | WAS | |
| With Determiners | 11221 | 9854 | 77.5% | 78.1% | |
| No Determiners | 3264 | 2769 | 22.5% | 21.9% | $\chi^2(1)=1.392$, p=0.238 |
| Total Nouns | 14485 | 12623 | 100% | 100% | |

## 5.5. Comparable Use of Determiners with Inherent Plurality

Question 5: Does WAS use more cardinal numbers or quantifiers than Castilian?

WAS does not use more cardinal numbers or quantifiers than Castilian. There is no

significant difference in token frequencies for FEM Quantifier+N phrases: $\chi^2(1)=2.858$,

p=0.090 (Table 20). Neither are there any significant differences in token frequencies for

masculine Quantifier+N phrases: $\chi2(1)=0.765$, p=0.381.

Table 20

*Usage of Quantifier+N Phrases in Castilian and WAS*

| | | Tokens | | |
|---|---|---|---|---|
| Quantifiers | Noun | CaS | WAS | Chi-square |
| FEM | C-final | 173 | 190 | |
| | V-final | 246 | 213 | $\chi^2(1)=2.858$, p=0.090 |
| MASC | C-final | 45 | 50 | |
| | V-final | 308 | 282 | $\chi^2(1)=0.765$, p=0.381 |
| No Gender | V-final | 13 | 15 | |

There are also no significant differences found in the use of cardinal numbers

between WAS and Castilian: $\chi2(1)=0.004$, p=0.950 (Table 21).

Table 21

*Usage of Cardinal+N Phrases in Castilian and WAS*

| Cardinals | Noun | Tokens | | Chi-square |
|-----------|------|--------|------|------------|
| | | CAST | WAS | |
| | C-final | 93 | 78 | |
| | V-final | 629 | 533 | $\chi^2(1)=0.004$, p=0.950 |

Additionally, there are no significant differences in the use of double determiners between WAS and Castilian: $\chi^2(1)=0.466$, p=0.495 (Table 22). These combinations involve a cardinal or a quantifier plus another determiner. The quantifiers and cardinals cue plurality inherently by way of their lexical meaning.

Table 22

*Comparing Determiner + Determiner + N Phrases*

| DET + DET | Castilian | Was | Chi-square |
|-----------|-----------|-----|------------|
| C-final | 70 | 61 | |
| V-final | 532 | 383 | $\chi^2(1)=0.466$, p=0.495 |

## 5.6. Summary

In this section, we have seen that there are no statistically significant lexical differences between these dialects, and only one significant morpho-syntactic difference in DET+N phrase frequencies, in the FEM INDEF ARTs. There is no evidence in this data of large-scale compensation for [s] lenition in WAS.

## 6. Discussion

Recall that the purpose of this thesis is to test the claim that [s] lenition results in a semantically critical loss of plural marking in Spanish, and therefore requires compensation. My results support the hypothesis that there is no semantic pressure for WAS speakers to compensate for word-final [s] lenition because word-final [s] has a very low semantic relevance in Spanish. There is no evidence of compensation in 98.7% of nouns compared between WAS and Castilian.

There is no evidence in this data that WAS and Castilian differ materially in lexical complexity or that WAS speakers compensate for [s] lenition by way of lexical restructuring. WAS speakers do not use more cardinal numbers or quantifiers than Castilian. WAS does not use more determiners than Castilian. Finally, only one statistically significant result is found, and this is in phrases of the construction FEM INDEF ART + NOUN.

The specificity of this phrase environment raises questions and may be evidence that [s] was a relevant contrast at an earlier stage of the language. The significant result occurs with *una/s.* If there were going to be statistically significant results that were due to a lack of number cues, then this is exactly the variable that would be expected to stand out. In WAS, due to the loss of [s], the singular and plural form of the FEM INDEF ART, *una* ("a" or "some") is homophonous with the cardinal number, *una* ("one"):

| | | | | |
|---|---|---|---|---|
| **(1)** | Castilian: | una amiga | = | a friend / one friend |
| | | unas amigas | = | some friends |
| | WAS: | una amiga | = | a friend / one friend / some friends |

The statistically significant reduction in use of the WAS FEM INDEF ART, *una*, compared to Castilian, could be due to the real potential for semantic ambiguity about "one" versus "some."

There is an alternate way to express "some" instead of *una*, by way of the quantifier *alguna* ("some"). I did find some use of the *alguna* forms in both WAS and Castilian, but only rarely, and there were not enough tokens for WAS speakers to be using *alguna* as a full replacement for *una*.

The less frequent use of *una/s* in WAS may be the one situation in which historical [s] lenition resulted in phonological neutralization (Gurevich 2004), such that there were semantic consequences for the historical loss of [s]. The different uses of *una/unas* at different times in Spanish history is documented in Pozas-Loya 2010. My suspicion is that certain FEM INDEF ART +N phrases proved to be particularly ambiguous in WAS, causing speakers to opt for using the bare noun instead, and that this tendency was passed along over time to emergent speakers, ultimately resulting in a slight dialect difference in this variable. Some evidence supporting this idea is found in Miller and Schmidt 2005 and 2012, which document that children in the variable [s] lenition dialect of Chilean Spanish almost never produce indefinite articles, preferring bare nouns (2012), whereas children from a non-lenition dialect in Mexico do consistently use una and unas. Miller and Schmidt also find that children from Chile interpret indefinite articles differently than bare nouns (2003). It may be that bare nouns play a specific role in [s] lenition dialects, but not in non-lenition dialects.

My study is not designed to capture that level of detail, however, and while there may be subtle differences in the use of bare nouns between WAS and Castilian, there is no statistically significant difference between these dialects in the broad frequency totals for all bare nouns (Table 19). Therefore, it is also possible that there is no compensation whatsoever, even in these doubly homophonous phrases, and that the statistically significant differences are due to random chance.

## 6.1. How Do Listeners Decode Plurality without [s]?

Ranson 1993 demonstrates that instead of [s], WAS speakers rely on other features of the Spanish language to provide number contrasts, such as verb inflection, determiners, and context. This can be seen in the data of the current study as well.

The corpora chosen for this study were both created in the same way. Both dialects use the same writing system, so the WAS transcribers wrote "s" on any nouns and determiners that they understood as plural. However, As Ranson 1993 notes, only a fraction of the written "s" sounds would have been produced in the WAS corpus (1.5% in Ranson 1993), and they would not be produced in a patterned way.

In (2), an example of a fragment from the Málaga corpus (WAS) is analyzed:

| **(2)** | *fue* | *a* | *la* | *tienda* | *a* | *comprarnos* | *unas* | *corbatas* |
|---------|-------|-----|------|----------|-----|--------------|--------|-----------|
| | he/she | to | the | store | | to buy us | a/some | tie(s) |
| | went | | DET.ART | NOUN | | | DET.ART. | Noun |
| | | | DEF.F.SG | F.SG. | | | INDF.F.PL | F.PL |

Probable gloss: He/she went to the store to buy us some ties.

*Tienda* is part of a homophonous DET+N phrase because both *la* and *tienda* are singular/plural homophones. Although the transcriber interpreted and marked *tienda* as singular, there would be no way to recover plural cues from within the phrase itself, so

the transcriber's assessment would be based on context. *Corbatas* is also a V-final noun, preceded by a feminine determiner, this time an indefinite article. In this particular fragment, the transcriber presumably understood *tienda* "store" as singular because if the person went to more than one store, a definite article would not have been used. *Corbatas* is marked as plural because it would not make sense for someone to buy one tie for more than one person to share, and the 1st person PL pronoun appended to the verb construction overtly shows that there are multiple people receiving ties. Therefore, the structural ambiguity of these nouns and determiners, caused by what would be a lack of [s] production in the spoken version of the corpus, does not lead to real semantic ambiguity.

The fragment, *esas naciones* "those nations" in (3), would be spoken as [esa nasione] in WAS, and although the demonstrative adjective would not carry number marking, the remnant [e] on the end of [nasion] signals plurality.

(3)  ***dentro   de   esas               naciones***
     inside    of   those-            nations-Noun.F.PL.
                    DET.DEM.ADJ.F.PL
     Probable gloss: "inside of those nations'

Throughout the transcripts of these corpora, fluent plural disambiguation can be seen in action. Speakers convey the information they wish to convey. The native speakers transcribing the interviews reveal their interpretation of grammatical number by whether they choose to represent a spoken noun with a word-final [s] or not. I was unable to find a single example in the WAS corpora of a listener misunderstanding grammatical number

such that it resulted in confusion. Tellingly, the rate of orthographic number-marking is comparable in each corpus.

**6.2. Why Might It Work this Way?**

It is possible that the expectation of compensation for [s] lenition was correct all along, but that the needed compensation already occurred diachronically, in the Middle and Golden Ages (roughly 1200-1600). If the increase of determiner use seen in Spanish literature, as documented by Lapesa (1973) from 1200 to the 1980s, mirrors the habits of speakers through time, then perhaps the disambiguating nature of the increased determiner use created an environment in both Castilian and WAS in which the potentially less salient final sibilant began to have less and less semantic pressure to resist lenition.

The initial period of increased determiner use seen in Lapesa 1973, takes place during same time period that [s] lenition was becoming established as a norm in Andalusia, which is thought to be from the 1400s to the 1600s (Penny 2002). The increased use of determiners may have co-evolved with the increase of lenition. The idea of compensation for Spanish [s] lenition as derived from diachronic morpho-syntactic changes in the Middle and Golden Ages, is inspired by explanations of similar circumstances of neutralization and syntactic change that may have co-evolved in Korean and Chinese, as discussed in Silverman 2010, 2012, and 2015.

This hypothetical scenario begs the question of why Castilian would have retained [s] if both dialects evolved to not rely on it as a plural contrast. One possible answer to this question is the strong and ongoing prescriptive influence on Castilian Spanish that dates

back to Alfonso X of Castile and Leon (1252-84), who is famous for the creation of an early standard of Spanish, and was known for his concern for "correctness" in language (Penny, 2002, p. 20). This early standardization in Castilian may still be driving the socio-linguistic use of [s] in Castilian to this day.

**6.3. Neutralization or Merge?**

Is it still correct to call this sound change neutralization, instead of a merge? Semantically, there are arguments to be made that this is still a synchronic sound change because people are still consciously aware of it. There is a stigma to not producing [s], and adults sometimes produce it, unlike the silent e of English, for instance, in which there is no social pressure to acknowledge this lost sound. Phonologically, on the other hand, this sound may be a diachronic merge. Once a sound change reaches the point at which children do not produce it, and they must memorize when to represent it orthographically in school, it becomes difficult to argue that the sound is in the phonetic inventory of the children. The fact that some adults may acquire word-final [s] later is not the same process.

## 7. Conclusion

This thesis proposes a new methodology for investigating Spanish [s] lenition that is not dialect-specific, and which does not require acoustic analysis or impressionistic coding. I demonstrate with quantitative data that WAS does not differ substantially from Castilian in the use of potentially ambiguous nouns, despite a demonstrable reduction of morphological plural marking. Additionally, I find that Spanish [s] lenition is a useful case study to show that neutralization can proceed, even if it results in large-scale derived homophony, provided that the lost contrast has a low semantic relevance. My data suggest that semantic relevance should be considered in discussions of functional load.

Castilian and WAS historically evolved side-by-side, and most modern dialects have their roots in one of these two dialects, therefore the comparable frequencies of potentially ambiguous nouns and phrases implies that the conditions that allow [s] lenition to propagate are inherent in the structure of the Spanish language itself, rather than specific to lenition dialects. Whatever the historical role of [s] as a plural marker, it appears that redundant number agreement, combined with a robust determiner system has rendered word-final [s] largely semantically irrelevant in Modern Spanish.

If [s] has no semantic pressure to resist lenition, then the manner in which it lenites simply becomes a question of accent, rather than function. This explanation works on a global scale to explain how word-final [s] lenition has been able to proceed in such variable and gradient ways, in so many different Spanish dialects, with no counter-functional consequences within those dialects. Word-final [s] does appear to be expendable for plural marking in Modern Spanish.

## References

Carlson, K. M. (2012). An acoustic and perceptual analysis of compensatory processes in vowels preceding deleted post-nuclear /s/ in Andalusian Spanish. *Concentric: Studies in Linguistics, 38*(1), 39-67.

Defior, S., Alegría, J., Titos, R., & Martos, F. (2008). Using morphology when spelling in a shallow orthographic system: The case of Spanish. *Cognitive Development*, *23*(1), 204-215.

Erker, D. G. (2010). A subsegmental approach to coda/s/weakening in Dominican Spanish. *International Journal of the Sociology of Language, 2010*(203), 9-26.

File-Muriel, R. J. (2010). Lexical frequency as a scalar variable in explaining variation. *The Canadian Journal of Linguistics/La Revue Canadienne De Linguistique, 55*(1), 1-25.

File-Muriel, R. J. (2012). A laboratory approach to s-lenition in the Spanish of Barranquilla, Colombia. *In Richard J. File-Muriel & Rafael Orozco (eds.), Colombian Varieties of Spanish, Madrid: Iberoamericana Vervuert. 127-140.*

File-Muriel, R. J., & Brown, E. K. (2011). The gradient nature of s-lenition in Caleño Spanish. *Language Variation and Change, 23*(02), 223-243.

Gurevich, N. (2004). *Lenition and contrast: The functional consequences of certain phonetically conditioned sound changes.* Psychology Press.

Hernandez-Campoy, J., & Trudgill, P. (2002). Functional compensation and southern peninsular Spanish/s/loss. *Folia Linguistica Historica, 36*(Historica vol. 23, 1-2), 31-58.

Labov, W. (1994). Principles of language change. Volume 1: Internal factors. Cambridge: MA: Blackwell

Lapesa, R. (1973). Un, una como articulo indefinido en espanol. In R. Cano Aguilar & M.T. Echenique Elizondo (EDS.), Estudios de morfosintaxis historica del espanol, (pp. 477-87). Madrid: Gredos

Málaga Urban Vernacular. (2007). *PRESEEA-MA Spoken Corpus* [Data files]. Retrieved from http://www.vum.uma.es/en/spoken_corpus.php

Marrero, V. and C. Aguirre. 2003. Plural acquisition and development in Spanish. *Linguistic Theory and Language Development in Hispanic Languages: Papers from the 5ᵗʰ Hispanic Linguistics Symposium*. Vol. 2, 275-296.

Martinet, A. (1952). Function, structure, and sound change. *Word-Journal of the International Linguistic Association, 8*(1), 1-32.

Miller, K. (2007). *Variable input and the acquisition of plurality in two varieties of Spanish* (Doctoral dissertation, Michigan State University).

Miller, K. & Schmitt, C. (2005). The interpretation of indefinites and bare singulars in Spanish child language. In *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, ed. David Eddington, 92-101. Somerville, MA: Cascadilla Proceedings Project

Miller, K., & Schmitt, C. (2010). Effects of variable input in the acquisition of plural in two dialects of Spanish. *Lingua, 120*(5), 1178-1193.

O'Neill, P. (2005). Utterance final/S/in Andalusian Spanish. The phonetic neutralization of a phonological contrast. *Language Design: Journal of Theoretical and Experimental Linguistics,* (7), 151-166.

O'Neill, P. (2009). S-Aspiration and occlusives in Andalusian Spanish: Phonetics or phonology?. *Oxford University Working Papers in Linguistics, Philology & Phonetics*, *12*, 73-85.

Parrell, B. (2012). The role of gestural phasing in Western Andalusian Spanish aspiration. *Journal of Phonetics, 40*(1), 37-45.

Penny, R. J. (2002). *A history of the Spanish language*. Cambridge University Press.

Poplack, S. (1980). The notion of the plural in Puerto Rican Spanish: Competing constraints on (s) deletion. *Locating Language in Time and Space, 1*, 55.

Pozas Loyo, J. (2010). The Development of the indefinite article in Medieval and Golden-Age Spanish, Queen Mary University of London, unpublished doctoral thesis.

PRESEEA. (2014). *Corpus del Proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. [http://preseea.linguas.net].

Ranson, D. L. (1993). The interaction of linguistic and contextual number markers in Andalusian Spanish. *Hispania, 76*(4), 919-930.

Romero, J. (1995). An articulatory view of historical /s/-aspiration in Spanish. *Rivista Di Linguistica, 7*, 191-208.

Ruch, H. (2013). Investigating a gradual metathesis: Phonetic and lexical factors on/s/-aspiration in Andalusian Spanish. *University of Pennsylvania Working Papers in Linguistics, 19*(2), 19.

Ruch, H., & Harrington, J. (2014). Synchronic and diachronic factors in the change from pre-aspiration to post-aspiration in Andalusian Spanish. *Journal of Phonetics, 45*, 12-25.

Sandoval, A. M., Toledano, D. T., de la Torre, R., Garrote, M., & Guirao, J. M. (2008). Developing a phonemic and syllabic frequency inventory for spontaneous spoken Castilian Spanish and their comparison to text-based inventories. In *Language Resource and Evaluation Conference* (pp. 1097-1100).

Silverman, D. (2010). Neutralization and anti-homophony in Korean. *Journal of Linguistics, 46*(02), 453-482.

Silverman, D. (2012). *Neutralization*. Cambridge University Press.

Silverman, D. (2015). The enigma variations: Response to Sampson. *Journal of Chinese Linguistics*, *43*(2), 697-702.

Silverman, D. D. (2006). *A critical introduction to phonology: Of sound, mind and body*. London; New York: Continuum.

Terrell, T. D. (1979). Final /s/ in Cuban Spanish. *Hispania*, *62*(4), 599-612.

Torreira, F. (2007). Pre-and postaspirated stops in Andalusian Spanish. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4, 282*, 67.

Torreira, F. (2012). Investigating the nature of aspirated stops in Western Andalusian Spanish. *Journal of the International Phonetic Association*, *42*(1), 49-63.

Torreira, F., & Ernestus, M. (2012). Weakening of intervocalic /s/ in the Nijmegen corpus of casual Spanish. *Phonetica, 69*(3), 124-148. doi:10.1159/000343635

Schmid, H. (1995). Treetagger, a language independent part-of-speech tagger. *Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart*, *43*, 28.

Wedel, A., Kaplan, A., & Jackson, S. (2013). High functional load inhibits phonological contrast loss: A corpus study. *Cognition, 128*(2), 179-186.