The Role of Social and Linguistic Factors in the Preservation of Arabic-Based Lexicon in Modern Spanish

By

Ryan Grosvold

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Spanish)

at the

UNIVERSITY OF WISCONSIN-MADISON

2021

Date of final oral examination: 05/03/2021

The dissertation is approved by the following members of the Final Oral Committee:

Fernando Tejedo-Herrero, Associate Professor, Spanish
Catherine Stafford, Associate Professor, Second Language Acquisition
Rajiv Rao, Associate Professor, Spanish
Grant Armstrong, Associate Professor, Spanish

Abstract

Hispanists have largely focused on the loss of Arabisms (i.e., Arabic loanwords) in Castilian while largely neglecting the reasons for their maintenance and incorporation in the language. With maintenance and incorporation of Arabisms in mind, the present study will address the following questions: (1) What specific factors allowed Arabisms to coexist with Romance terms with which they are synonymous?, (2) What processes supported retention of old Arabisms and introduction of new Arabisms in the language at the same time that other Arabisms were ousted from the Spanish language?, and (3) More generally, why were certain Arabisms maintained or introduced when others were not? The study is important for the field of historical Spanish linguistics in that it addresses the treatment of lexical rivalries of words that are semantically related but are either (1) Latinisms, patrimonial terms, or Romance borrowings or (2) Arabisms. In addition to providing more complete histories of three sets of lexical rivalries (i.e., *aceite-olio-óleo*, *atalaya-centinela,* and *jaqueca-hemicránea-migraña*), the dissertation will identify and analyze the conditions that were favorable to the maintenance and incorporation of Arabisms in a linguistic environment that was largely anti-Semitic. The potential significant factors that we have identified in the maintenance and incorporation of Arabisms in Spanish are the following: (1) formal variation, (2) related vocabulary, (3) semantic differences, (4) dialect, (5) register, (6) gender, (7) age, and (8) circumstances of language contact.

Table of Contents

List of Tables

1 Introduction

1.1 Historical background

At the beginning of the eighth century, the Moorish civilization reached the West, converting areas of Roman territory into a *Romania Arabica* that consisted of the Iberian Peninsula, Malta, and Sicily. In this territory there lived peoples of various religions and cultures, including Christians, Jews, Muslims, Arabs, Visigoths, and Berbers. While the Moors also exerted their influence on Sicily and Malta, the Iberian Peninsula is the only territory that was under Moorish occupation (at least in part) for nearly 800 years between 711 and 1492 A.D., during which time the various cultures and religions coexisted in the Peninsula, exercising mutual cultural influences on one another. Although the Christians had reclaimed the Peninsula in its entirety by 1492, various Arabic-speaking populations remained in Iberia for centuries thereafter. In this confluence of Peninsular cultures, the incorporation and elimination of certain aspects were not limited to social or cultural practices but also applied to the linguistic sphere. In addition to developments in other Ibero-Romance languages (e.g., Catalan and Portuguese), the linguistic contribution of Arabic to Castilian is considerable, both in number of words and in variety (Winet 2006).

Among all Romance languages, the Ibero-Romance languages are those that experienced the greatest influence from Arabic. While Portuguese and Catalan have maintained many words of Arabic origin, Castilian has maintained the greatest number of words of Arabic origin of all extant Romance languages. The reach of Arabic influence in Spanish is so considerable that, apart from Latin, Arabic is the source of the greatest number of borrowings in Spanish (Colón

1998, Corriente 2008). This Arabic element, in reference to the lexical, semantic, and syntactic influences in Romance, falls under the category of Arabisms. Although some researchers have analyzed the influence of Arabic on Romance syntax and semantics, most have focused their attention on lexical borrowings, the most obvious and substantial influence from Arabic in this language group (Winet 2006). For this reason, I will limit my study to lexical influence.

Among the semantic categories of the Spanish lexicon expanded through contact with Arabic, the following are some of the more common ones: knowledge and science (*alcohol*, *cifra*), construction (*alcázar*, *azulejo*), urbanism and housing (*arrabal*, *barrio*), domestic life (*almohada*, *zapato*), agriculture (*acequia*, *aljibe*), culinary terms, plants, flowers, and fruits (*aceituna*, *albaricoque*, *alcachofa*, *algodón*, *arroz*, *naranja*), animals (*alazán*, *atún*), the military (*algara*, *atalaya*), mineral products (*azogue*, *azufre*), musical terms (*guitarra*, *tambor*), and colors (*azul*, *escarlata*) (Giménez-Eguíbar 2011). It should not be surprising, then, that this distribution of nouns corresponds to the sociocultural aspects that experienced the greatest influence from the Moorish presence in the Peninsula (Lapesa 1981).

Despite their variety and abundance in the Middle Ages and beyond, over time disdain toward Arabisms grew significantly. This contempt for the Arabic influence in Castilian emerged most notably in the early modern period, when speakers of normative varieties began to concern themselves with the standardization of Spanish. Authors of works of this kind shared the notion that any incorporation of words of Arabic origin in Castilian was to the detriment of the language. Since standard Castilian is an idealized construction that interpreted the Arabic influence as invasive, grammarians sought to eliminate them through the incorporation of

Latinisms. It appears that such linguistic attitudes indicated that Arabisms hardly formed part of the standard, that is, that they did not belong in the language, especially that of the educated class (Giménez-Eguíbar 2011).

As a result of the negative attitudes that existed, many Arabisms of frequent usage fell into various levels of disuse, though many are still common today. Three levels of use and disuse of Arabisms can be identified: An Arabism (1) may be the principal term for a particular meaning (*azul* 'blue'), (2) may coexist with a Latin term, typically as a regional or rustic variant (*alhucema* 'lavender' vs *lavanda* 'lavender (plant); lavender perfume; lavender (color)'), or (3) may have truly been lost since there is no written record of it (Gilman 1979).

Arabisms that became obsolete experienced many changes, such as semantic restriction, semantic extension, and the restriction in the number of meanings of a given term. Since the process is gradual, it required a different amount of time for each word since, for a time, lexical variants coexisted, and the choice of one variant or another varied depending on the geographical region in which they were used (Giménez-Eguíbar 2011).

From their appearance in the Middle Ages to their eventual decline, Arabisms have played a significant role in the development of the Castilian lexicon. While their gradual exclusion and replacement are now well accounted for, other questions remain unanswered. For example, even a study as comprehensive as that of Winet (2006) concluded that there is no single determining factor for the retention (or elimination) of the Arabic article in Spanish words of Arabic origin. Instead, the author claims that the phenomenon is determined by a complex

combination of intralinguistic and extralinguistic factors. We also believe that we will find that no Arabism is maintained due to any single favorable linguistic or sociolinguistic factor alone. Instead, we expect that retention may be favored (or disfavored, as the case may be) by several different factors.

1.2 Purpose of the present study

Ample written evidence reveals the importance that the Arabic language exercised on medieval and early modern Iberia. This influence has been displayed in numerous diachronic accounts of the Spanish language and lexicon, including Neuvonen (1941), Lapesa (1981: 129-156), Maíllo Salgado (1998), Corriente (2005, 2008), Dworkin (2012), and other authors. Two important linguistic events that represent the lexical transition of Castilian during the Middle Ages and the early modern period have gained some attention among Hispanists: (1) the introduction of learned terminology, whether of Latin, Greek, or Hebrew origin (Lapesa 1981) and (2) the replacement of Arabisms that had previously enjoyed vitality (Eberenz 2006: 85-102). Still, these two phenomena were only scarcely studied to the degree that Eberenz noted how researchers have tended to avoid lexical rivalries between Arabic terms, on one hand, and Latin or Romance terms, on the other:

> Aunque cada vez sepamos más sobre los arabismos del español, carecemos de estudios lexicológicos sobre su relativa decadencia o, más concretamente sobre la rivalidad entre ciertos arabismos y sus equivalentes de origen europeo, como *alfajeme* y *barbero*, *alfayate* y *sastre*, *almojarife* y *recaudador*, *albóndiga* y *depósito real* y *campamento de tropa*, *azogue* y *mercurio*, *almoneda* y *subasta*. Frente a la corriente latinizante, tan importante para la renovación del español, el retroceso de la influencia oriental y el progresivo confinamiento de los arabismos a la esfera rústica están aún mal estudiados… Los tratados científicos plantean, además, otro problema estrechamente relacionado con nuestro tema: me refiero a los arabismos y a su paulatina sustitución por elementos de procedencia distintas… (Eberenz 2006: 98-99)

In terms of the state of the Castilian lexicon, the fifteenth century is highly significant since it is the period when Arabisms nearly ceased to enter into Castilian language. While Corriente (2008) implies that Arabisms no longer entered into the Spanish language after the fifteenth century, research has found that this assertion is not entirely correct. While it is true that, in the fifteenth century and beyond, Arabisms no longer entered into Castilian in the large numbers as they once had in the medieval period, there are in fact several additional entries of Arabic origin into Spanish during the early modern period, on the one hand, and the twentieth century (Morgenthaler García 2014), on the other.

Considering the imposed restriction of Arabisms from entering into the Spanish language in the fifteenth century, one cannot deny that this fact is largely due to the incipient institutionalized rejection of words of Arabic origin during the early modern period. Documenting this rejection, which eventually led to the loss of Arabisms from Castilian language, Giménez-Eguíbar (2011) addresses the lack of studies of Arabic-Romance lexical rivalries in her study of the semantic field of professions. Her thorough study highlights the process of lexical loss to account for the elimination of certain Arabisms from Castilian language.

While Giménez-Eguíbar and other researchers have focused on the more conspicuous elimination of Arabisms from Castilian, there still remains a lack of research in the area of maintenance and incorporation of Arabisms. Even though the status quo was the elimination of the Arabism from Castilian, especially in the presence of a Romance or Latinate synonym, there are numerous cases in which Arabisms in competition with Romance terms did survive. Given

this state of the language, a few questions must be raised: (1) What specific factors allowed Arabisms to coexist with Romance terms with which they are synonymous?, (2) What processes supported retention of old Arabisms and introduction of new Arabisms in the language at the same time that other Arabisms were ousted from Castilian?, and (3) More generally, why were certain Arabisms maintained or introduced when others were not?

In other words, the goal of the present investigation is to provide a more complete history of the specific Arabisms considered and to provide explanations for maintenance and incorporation of Arabisms in spite of a harsh environment of linguistic prescription and negative sentiment that otherwise would have eliminated them entirely from the language. The study is by no means exhaustive in its selection of Arabisms that are incorporated into and maintained in Castilian; however, the lexical items chosen are intended to be representative of the greater selection of Arabisms that fall under one or another category. In addition, parallels will be drawn from other languages, as appropriate, to determine the behavior of Arabisms in different linguistic environments in order to support claims that are relevant for Arabisms in Castilian.

1.3 Arabisms, defined

For the purposes of this study, the term *Arabism* will refer to a word borrowed from Classical Arabic or Hispano-Arabic into Castilian. We make one further distinction for our study, that of direct and indirect Arabisms, based on their transmission. We have identified Arabisms entering into Castilian through speech as *direct Arabisms* and those introduced through writing as *indirect Arabisms*, acknowledging that different authors apply these terms in different ways.

Other authors offer other classifications of Arabisms. For example, Winet (2006) categorizes Arabic loanwords as false, uncertain, or certain Arabisms. False Arabisms are words that appear to be (but are not) Arabisms based on their form (e.g., Sp. *almendra* 'almond' < Vulg. Lat. *AMYNDULA), whereas uncertain Arabisms are words that have etyma that cannot be confirmed (e.g., Ext. *aldobara* 'rivulet' < Hispano-Ar. *addawwára* 'circle, disk'?). The only one of these categories that will be of interest for the present study are certain Arabisms, which have etyma that are confirmed by specialists (e.g., Corriente 2008).

Perhaps the most obvious feature of an Arabism in Castilian is the presence of the Arabic article *al-* or *a-* (the latter of these having resulted through assimilation in point of articulation), which made it easily identifiable to grammarians who would later denounce the use of Arabisms by simple identification of the prefix. Of course, incorrect identification of non-Arabisms as Arabisms did occur, especially through misidentification of initial *al-* as Arabic in origin (e.g., *almendra* 'almond' < Vulg. Lat. *AMYNDULA) (Winet 2006: 275). The Arabic article will be relevant as we examine the circumstances of language contact since speaker awareness should be greater in the presence of this morphological feature in an Arabism.

1.4 Hypotheses

From their appearance in the Middle Ages to their eventual decline, Arabisms have played a significant role in the development of the Castilian lexicon. Although the gradual exclusion and replacement of Arabisms have received attention in the literature, we must still address why they were maintained.

While reviewing previous studies, we realized that several of the internal and external factors analyzed in sociolinguistic research could be analyzed for a diachronic analysis of Arabisms.[1] After compiling a large number of Arabisms with Romance or Latinate counterparts, we determined that the factors most likely to contribute to retention were semantic differences, attitudes, and transmission, prompting the following hypotheses:

(1) An Arabism is maintained as a partial synonym

We expect that an Arabism with a Latinate or Romance counterpart is more likely preserved if it is a partial (or functional) synonym of the Latinate or Romance term, rather than a true synonym,[2] whether the differences in meaning are the result of semantic changes or those that had existed since its introduction. If an Arabism undergoes semantic restriction or generalization, it establishes a contrast in meaning with the word with which it was once, but no longer, closely synonymous. This claim applies to all three case studies in the present study (i.e., *aceite*, *atalaya*, *jaqueca*), albeit for different reasons. OSp. *azeyte* 'olive oil' > MSp. *aceite* 'oil' might never have been a true synonym of OSp. *olio* 'oil' or MSp. *óleo* 'plant-based oil; oil paint; oil painting; holy oil.' *Atalaya* 'sentinel; watchtower' > 'watchtower', on the other hand, initially had a distinctive meaning that its counterpart *centinela* 'sentinel' never had, and in competition the former term lost the meaning that it had shared with the latter. Finally, although it is not strictly a case of semantic differences, *jaqueca* and its original counterpart, *hemicránea*, have always maintained distinctive register usage.

(2) An Arabism is maintained with neutral or favorable attitudes

[1] Previous research of internal and external factors is reviewed in sections 2.3.3 and 2.3.4, respectively.
[2] We discuss the differences between partial (or functional) synonyms and true synonyms in section 2.3.2.

We also propose that an Arabism is maintained if speaker attitudes toward a given Arabism (or domain) are neutral, as opposed to the generally negative attitudes that Arabisms faced, especially in the early modern period. This factor applies to all three case studies as well. In other words, *aceite*, *atalaya*, and *jaqueca* must have been viewed at least neutrally, if not favorably, in the course of their usage.

(3) Learned, or indirect, Arabisms should be better retained than popular, or direct, Arabisms

Since indirect Arabisms are introduced through writing and, therefore, associated with higher registers, they should be more resistant to elimination than Arabisms transmitted through speech, belonging to the popular register.[3] This claim may be evaluated through a comparison of the degree of retention of scientific Arabisms, on one hand, and agricultural and military Arabisms, on the other. We also know that *jaqueca* and *jarabe* were retained while others were not, which means that we must determine their transmission (i.e., speech or writing) in order to compare them with the scientific Arabisms that were eliminated.

1.5 Structure of the study

This study is divided into six chapters, of which the first offers a general introduction to Arabisms in Castilian. Above, we have provided a brief historical overview of the contact between Arabic and Castilian Romance speakers, the purpose of our study, our definition of Arabisms, and our hypotheses.

[3] We will explain our classification of direct and indirect Arabisms further in the following chapter, in section 2.3.4.3.3.

Chapter two discusses our theoretical approach to Arabisms, including methodological and theoretical difficulties associated with research in lexical variation in general, as well as how these apply to studies that examine Arabisms in particular. We also examine the roles of linguistic and sociolinguistic factors in lexical variation and, in particular, how these have the potential to explain the retention of Arabisms. Finally, we explain our methodology in approaching our data concerning the Arabisms (i.e., *aceite*, *atalaya*, *jaqueca*) and semantic domains (i.e., agriculture, the military, medicine) in question.

In chapter three, we examine *aceite* as a representative of the broader domain of agriculture. We argue that its introduction into Castilian is the result of a perceived need to identify a new material reality and that its retention is due, in large part, to semantic differences that it maintained with its Romance counterparts, as well as neutral attitudes toward the Arabism. Since many of the linguistic and sociolinguistic factors that favor *aceite* are also those that favor other agricultural Arabisms, *aceite* may be considered a typical agricultural Arabism.

In chapter four, we analyze the treatment of *atalaya* as a military Arabism. We contend that the Arabism is introduced into Castilian Romance due to the infrequency (or absence) of equivalent patrimonial terms and that its maintenance is motivated by semantic differences between *atalaya* and *centinela*, and also neutral attitudes toward this Arabism and other, similar Arabisms. *Atalaya* may be considered a typical military Arabism since the linguistic and social factors that favor *atalaya* also favor other military Arabisms.

In chapter five, we address the treatment of *jaqueca* as a medical Arabism. We claim that this Arabism is introduced in the absence of an equivalent term for 'migraine' and that the term is retained due to differences in register between it and its Latinate counterpart, as well as neutral attitudes toward the Arabism. Based on the linguistic and sociolinguistic factors that favor *jaqueca* and its retention, we do not consider *jaqueca* to be a typical medical Arabism, most of which are transmitted through writing and are usually eliminated.

In chapter six, our conclusion, we offer observations on our findings from the three case studies and explain how these relate to and differ from one another. Specifically, we address which linguistic and sociolinguistic factors most favor the maintenance of Arabisms in agriculture, the military, and medicine individually, as well as the factors that most favor Arabisms across the three disciplines. Analyzing numerous examples across the different semantic domains of Arabisms, our data offer support for our claim that semantic differences and the circumstances of language contact (i.e., neutral attitudes, intensity of contact) are the factors that most favor the retention of Arabisms.

2 The study of lexical variation in general and as it pertains to Arabisms

2.1 Sociolinguistic research and language variation

The field of sociolinguistics arose from the various disciplines of dialectology, contact linguistics, historical linguistics, sociology, and psychology (Koerner 1991: 65). It is for this reason that it is difficult to provide a single definition for sociolinguistics; still, in a broad sense, it is a discipline that examines the effects of culture and society on language, which is necessarily social in nature since language is not possible without those who produce it. As a variable medium, language reveals insight into the identity of an individual whenever he or she speaks, including his or her geographical origin, economic, and sociocultural information (Tagliamonte 2006: 3). While some areas of sociolinguistics emphasize certain aspects of language over others, suffice it to say that sociolinguistics is the study of language in use.

Even before the existence of sociolinguistics as a formal discipline, scholars had been aware of the existence of linguistic variation. What began as a casual acknowledgement of variation in language toward the end of the nineteenth century (e.g., Schuchardt) eventually became a dismissal of the novelty of such a claim, as Sapir suggests: “Everyone knows that language is variable” (1921: 147). Naturally, interest in language variation led researchers to seek explanations for its existence. One group of linguists, as in Fries and Pike (1949), proposed that variation could be explained by the fact that speakers have access to distinct coexisting linguistic systems and may alternate between one system or another according to different circumstances. However, the fact that speakers can combine features from different registers (e.g., informal conversation, formal writing) or varieties (e.g., standard speech, local variety)

invalidates such a position. Later research supported the idea that not only are speakers able to combine features from different registers or varieties, but that they do so quite frequently. Studies such as Weinreich, Labov and Herzog (1968), for example, have shown that speaker behavior demonstrates that there is a single linguistic system in which variants at different levels of language (e.g., lexicon, morphology, phonology, etc.) coexist.

Since its inception, sociolinguistic research has branched into different divisions, of which the most general is between sociolinguistics and sociology of language. As stated above, sociolinguistics proper is the study of language in its social context, while sociology of language studies the social implications of language. Naturally, then, any study of language variation belongs firmly within the discipline of sociolinguistics (Tagliamonte 2006: 3-4).

A subdiscipline of sociolinguistics that has garnered attention in recent years is historical sociolinguistics, which combines methodology from traditional historical linguistics and modern sociolinguistics. Whereas traditional historical linguistics supposes that languages change (giving priority to language-internal factors), modern sociolinguistics assumes that it is the speakers themselves who change language. In subscribing to the uniformitarian principle, then, whatever we discover about language variation and change today may also be applied to language variation and change in the past (Tuten and Tejedo-Herrero 2011: 285).

Historical sociolinguistics originated from and shares many of the goals of the subdisciplines of interactional sociolinguistics, sociology of language, and variationist linguistics. Sociology of language is concerned with language use and its status in society. As it

relates to Spanish historical sociolinguistics, some currents areas of interest are bilingualism, diglossia, and standardization (Moreno Fernández 2005). Interactional sociolinguistics typically uses qualitative discourse analysis to determine how speakers use language in context, with recent areas of interest in the pragmatics of Spanish texts, including analyses of register (Oesterreicher 2004) and genre (Kabatek 2008). The most influential contributor to historical sociolinguistics, however, is variationist sociolinguistics. As we already know, variationist sociolinguistics uses quantitative methodology to analyze the role of both internal and external factors in the use of a given variable and to identify language changes. Likewise, historical sociolinguistics aims to explain variation and change but often over large periods of time and with resources to which we have limited or no access. With this in mind, instead of analyzing considerable amounts of statistical data, historical sociolinguists tend to opt for careful examination of the language contained in historical texts (Tuten and Tejedo-Herrero 2011: 285-286).

Of all the subdisciplines of sociolinguistics, variationist sociolinguistics is perhaps the most influential and significant, due to its abundance in contemporary research. In a broad sense, variationist research is centered around the quantitative study of language variation. As a research tradition that analyzes both internal and external features of language, variationist sociolinguistics may be defined as "the study of the interplay between variation, social meaning and the evolution and development of the linguistic system itself" (Tagliamonte 2006: 4-5). Research in this subdiscipline relies on three features of natural language: (1) orderly heterogeneity, (2) constant change, and (3) its ability to convey non-linguistic information (Tagliamonte 2006: 5-6).

Orderly heterogeneity acknowledges that language varies from speaker to speaker. This applies to language on any scale, from differences between languages (e.g., Spanish, Arabic) to individual word choice in a single language (e.g., Sp. *alubia*, *frijol*). Another fact of language is that it is always changing, which comes as no surprise to anyone who has read literature in his or her own language from an earlier time period. Finally, language transmits information beyond the meaning of its individual words precisely because it varies. The kinds of information conveyed by an individual through language include his or her relationship with the interlocutor, social identity, age, gender, and socioeconomic status (Tagliamonte 2006: 6-7).

In addition to the three aspects of language (i.e., orderly heterogeneity, constant change, ability to convey non-linguistic information), another fundamental concept in variationist sociolinguistics is the linguistic variable. In section 2.2 below, we will define the linguistic variable and explain how its application has evolved. In sections 2.3.1 and 2.3.2 we will discuss the challenges of applying this concept to both synchronic and diachronic studies of lexical variation.

2.2 Language variation and the linguistic variable

While sociolinguists were concerned with the study of variation in language, generative linguists, under the concept of free variation, once believed that linguistic variability was unpredictable, random and, therefore, unworthy of the scientific discipline of linguistics. However, variationist research has demonstrated time and again that free variation does not

adequately explain how variants are often associated with certain linguistic and social factors that explain their usage (Blas Arroyo 2012: 28).

Confirming the value of linguistic variation as a formal discipline, Labov (1966) defines a linguistic variable as a variable structural unit, which is continuous and quantitative in nature. The linguistic variable is, as its name suggests, variable in the sense that it takes on different realizations under different circumstances, whether these are stylistic or social. The linguistic variable is continuous in that its variants often gain social value based on their proximity or distance from the standard variant. Finally, the linguistic variable is quantitative in nature since a given social value may be determined by the relative frequency of its variants.

As a variable unit, the linguistic variable is really "two or more ways of saying the same thing" (Labov 1972). Due to the transparency of variants at the level of phonology, then, it is no wonder that phonological variation has easily been the most popular area of sociolinguistic research. However, in recent years, scholars have shown a growing interest in the other levels, or types, of variation, as seen in the examples below, taken from the *Corpus sociolingüístico de Castellón* (Blas Arroyo 2009a):[4]

1. a. Puede ser que por él, a mí me hayan *gustado* las matemáticas (101)
   b. …una vez has *terminao* de estudiar… (101)
2. a. … algún trabajo, algo que le *fuera* a servir para él poderse ganar la vida (144)
   b. …si vivían durante el año donde *fuese*… (144)

[4] Each example set was taken from a different participant from the study, all of whom were coded numerically.

3. a. …este año hacen una fiesta para anunciar que *van a subir*… (198)

   b. es una fiesta muy muy bonita y este año, pues *subiremos* (198)

4. a. Vas a un campo y ves y dices: «¡ay esto, qué *bonito*!»… (158)

   b. …tengo un gato. ¡*Mono*, más *mono* él! (158)

In examples (1a) and (1b), we see variation of a phonological variable, in this case the phonetic realization of intervocalic /d/, specifically in cases of participle /ado/, a context subject to substantial variability in Spanish (Blas Arroyo 2012: 29-31). Since research began in sociolinguistics, phonological variation has been at the forefront since it is very frequent in speech and also because it is widely accepted that phonological variants are truly equivalent in meaning despite their differences of realization (Tagliamonte 2006: 70-71). The same may also be said of morphological variation, as seen in examples (2a) and (2b) where we see two variants of the past subjunctive (i.e., forms in *-ra* or *-se*). Although there is somewhat less research on morphological variation by comparison, researchers have found it possible to analyze such variation with few methodological concerns (see below for further explanation).

On the other hand, not all researchers would agree that syntactic variation, as found in examples (3a) and (3b), may be studied in the same manner as phonological or morphological variation. The reason for this is that some scholars, such as Lavandera (1978), suggest that the use of one syntactic variant or another signals a difference in meaning, which is the reason why they cannot be seen as variants of the same variable, given that the variants are not true equivalents. Still, there have been studies, such as that of Blas Arroyo (2008), which shows

syntactic variation appearing in the speech of a single speaker but also that the two variants identified (i.e., the analytic future and the synthetic future) are present in all contexts (120-121).

While syntactic variation already has its share of methodological difficulties, these are even greater in the case of lexical variation, which is why many researchers have tended to avoid analyzing it in detail. Despite the obvious synonymy of many sets of words (e.g., *bonito* and *mono*), as in examples (4a) and (4b) above, social and stylistic considerations in lexical variation analysis have made studies of this type of variation somewhat rare. The absence of literature on lexical variation is hardly limited to Spanish or English but is typical of most languages, despite the fact that research is quite extensive in both of these two languages with regard to the other levels of variation. Due to its methodological problems, mention of lexical variation is rare, even in large compendia dedicated to sociolinguistics and language variation. Reasons for the avoidance of research on lexical variation are many, with methodological issues and lower frequency of use being the most obvious (Blas Arroyo 2009b: 190-191). Another factor is criticism of the use of the linguistic variable for purposes other than the study of phonological variation.

Since the linguistic variable was initially and primarily intended for analyses of phonological variation, its application to other levels of linguistic analysis (e.g., morphology, syntax, lexicon) has been met with some opposition in the sociolinguistic research community. One notable exchange over this proposal began with Lavandera (1978), where the author declared that it was not possible for the sociolinguistic variable, initially developed as a tool for phonological analysis, to be used in analyses of morphological, syntactic, or lexical variation.

Citing two other studies, Sankoff (1972) and Laberge (1978), in which the selected variables showed no indication of social significance but rather syntactic constraints and a change in progress, respectively, Lavandera (1978) questions if the sociolinguistic variable is truly an appropriate tool for syntactic variation.[5] While the author recognizes the contributions of the linguistic variable in phonology, (e.g., establishment of the fact that variation is not simply free and that most differences in form signal a difference in meaning), she challenges the use of variable frequencies to analyze cases where speakers find themselves in a linguistic or social context that strongly influences their speech. For example, considering both standard *exhausted* and colloquial *wiped out*, the author states that a higher frequency of *exhausted* in formal contexts and a lower frequency of *wiped out* in informal contexts is not only is expected but that it also is not significant for a study of sociolinguistic variation. Lavandera explains that the difference in frequencies of these two terms is due to the fact that the forms themselves have differences in meaning. Furthermore, the author later questions if semantic equivalence is truly possible for variables in syntax, morphology, or the lexicon since morphemes, structures, and words all have meanings whereas phonemes do not (1978: 171-181).

In response to Lavandera (1978), Labov (1978) explains that while formal linguists are concerned with determining subtle differences in representational meaning (e.g., Eng. *was* vs *got* passive), sociolinguists recognize that morphemes, structures, or words with alike (but not identical) representational meaning may serve as variants of the same variable. Citing Sankoff and Thibault (1977), in which French auxiliary verbs *avoir* and *être* are determined to be

[5] The examples addressed in Sankoff (1972) are (1) placement of future marker, *bai*, in Tok Pisin and (2) elimination of complementizer *que* in Montreal French. The third example, from Laberge (1978), is placement of indefinite pronoun *on* in Montreal French.

equivalent in meaning in French past tense *passé composé*, Labov claims that for the purposes of sociolinguistic analysis, variants need not be exactly equivalent in meaning in all contexts. For this reason, it is essential that sociolinguists eliminate all contexts where the variants under consideration do not have the same representational meaning. Returning to the example of *exhausted* and *wiped out*, Labov admits that there are no absolute synonyms, but he also explains that style allows for a degree of flexibility in verbal expression so that words with different representational meanings may act as stylistic variants of one another. In cases of stylistic variation, two lexical items with otherwise different representational meanings may be treated as variants, provided they are intended to convey the same meaning (e.g., *trip, vacation*) (1978: 7-13). In other words, the linguistic variable can be extended to other levels of linguistic analysis, even if the task of circumscribing the variable context is less obvious, as in the case of lexical variation.

While the argument of Labov (1978) is useful because it can account for variation at any level of linguistic analysis, the study of lexical variation still has its share of challenges. Below we will discuss in more detail some of the methodological (section 2.3.1) and theoretical (section 2.3.2) challenges associated with lexical variation research.

2.3 Lexical variation as a general study and as it relates to Arabisms

In each remaining section of this chapter, we have organized the material in the same general format, first addressing general concerns and findings in lexical variation research before entering into a discussion of how these relate to Arabisms in particular. Sections 2.3.1 and 2.3.2 will address methodological and theoretical concerns. Later, in sections 2.3.3 and 2.3.4, we will

address linguistic and extralinguistic factors relevant in the study of lexical variation and consider their relevance to Spanish vocabulary of Arabic origin. The final section of this chapter, section 2.4, is dedicated to an explanation of our approach to studying the preservation of Arabisms in Spanish, the principal concern of the following chapters.

2.3.1 Methodological concerns in research on lexical variation in general and with regard to Arabisms

As stated earlier, one of the central issues in the study of lexical variation is that it is much rarer than phonological, morphological, or syntactic variation. While substantial data on phonological variation can be observed in a short amount of recorded speech, finding sufficient linguistic variation at the other levels requires a significant increase in time dedicated to data retrieval and analysis, especially in the case of vocabulary. That is, even if the desired lexical items do appear in variation, it is only possible to perform a successful analysis when there is adequate data to support it.

One strategy to address the issue of limitations of the amount of data available for lexical variation, and one which has long been in use in traditional dialectology and linguistic geography, is the questionnaire. Now also in use in contemporary sociolinguistics, the questionnaire has several advantages over the more traditional sociolinguistic interview: it allows the researcher to control more directly the kind of information collected and allows for a considerably greater amount of relevant data with a smaller investment of time and effort. Of course, there are different methods to elicit lexical variants in a questionnaire. One of the more common methods of elicitation is to offer the definition of a concept with the expectation that the

participant respond with one of the available variants in the target language. This is precisely the method employed in Almeida (1994), where the researcher analyzed variation of a set of taboo expressions and their respective euphemisms in a speech community in the Canary Islands.

While there are certainly benefits to using the questionnaire, it is not without its disadvantages. The first of these disadvantages is the fact that there is a certain degree of formality (or artificiality) in the speech of the informant with this type of data collection (García Mouton 1996: 76). That is, the main concern with regard to the questionnaire is that the participant is unlikely to produce spontaneous or vernacular speech, which are considered optimal for analyzing sociolinguistic variation (Labov 1984). Even under ideal conditions, a traditional sociolinguistic interview is not likely to elicit speech as authentic as that found under natural circumstances; language elicited by a questionnaire, then, is even less like natural speech than language elicited in a sociolinguistic interview. Accordingly, an analysis may be inaccurate if a participant chooses the variant that first comes to mind or avoids certain stigmatized variants instead of choosing his or her preferred variant (Borrego Nieto 1994). Even when a questionnaire is well planned and adequately accounts for the analysis of a stylistic continuum, there is always the possibility of overrepresentation of a given variant in contrast to its actual usage in the speech community or, on the other hand, the possibility that the researcher has selected variants that are not true synonyms of one another, in which case they cannot be considered variants since they do not have the same referent (Boberg 2005).

In an attempt to avoid the disadvantages that arise as a result of the use of a questionnaire, several researchers have eliminated steps in which the informant is not directly

asked questions, a strategy which certainly better approximates natural speech. For example, Geeraerts, Grondelaers and Bakema (1994), in their analysis of clothing items in Dutch, gathered data from thousands of images taken from magazines, a strategy that allowed them to determine several relevant factors, such as how geography influenced variation. Other studies have analyzed other types of written materials, such as corpora of social media, where participant language closely resembles natural speech (Grondelaers, Geeraerts and Tummers 2001).

Whenever possible, variationist sociolinguists analyze data from actual speech, often elicited in an interview. That is not to say that these materials are free of their own issues; as we addressed earlier, there is always the issue, in the study of lexical variation, of the low frequency of the targeted variable in real speech. Still, with the aid of sufficiently large corpora, the study of lexical variation of real speech is nevertheless possible.

While the above strategies are advantageous in a synchronic analysis of lexical variation found in speech, they are not helpful in a diachronic analysis of lexical variation concerned with written language. One solution to analyzing lexical variation in written language over time is to combine strategies from both traditional historical linguistics and modern sociolinguistics, an approach in line with Janda and Joseph, who explain:

> ...what we should really strive for, in diachronic pursuits such as historical linguistics, is what could be called *"informational maximalism"* – that is, the utilization of all reasonable means to extend our knowledge of what might have been going on in the past, even though it is not directly observable. Normally, this will involve a heavy concentration on the immediate present, but it is in fact more realistic just to say that we wish to gain a maximum of information from a maximum of potential sources: different times and different places – and, in the case of language, also different regional and social dialects, different contexts, different styles, different topics, and so on and so forth (2003: 37).

This hybrid approach, or what is properly known as the discipline of historical sociolinguistics, has allowed researchers to quantitatively analyze (written) language variation over time.

Historical sociolinguistic research has several advantages over synchronic studies on contemporary language. Unlike traditional sociolinguistic researchers, historical linguists generally do not influence the data that they acquire (i.e., the observer's paradox). Also, historical researchers are aware of the results of a given change in progress and can observe them over long periods of time; in contrast, contemporary language researchers, who are unaware of the results of a given change in progress, may only study them over relatively short periods of time. Furthermore, information that historical scholars have gathered from changes in earlier states of languages may help us to understand a present-day change in progress. Lastly, due to sociocultural differences between the past and present, historical research allows scholars to test hypotheses believed to be applicable to any time period (Tuten and Tejedo-Herrero 2011: 289).

Despite the possibilities of modern historical sociolinguistic research, the field is not without its challenges. Working with limited documentation and without access to additional data from the language users in question, historical sociolinguists do not have the same advantages as traditional sociolinguists researching contemporary language communities. Not only are historical sociolinguists often restricted to analyzing written language, but they must do so within a restricted variety of writing styles, genres, and participants (Tuten and Tejedo-Herrero 2011: 288). In other words, the vast majority of works studied in historical

sociolinguistic research are nearly always formal or academic in register and created by educated men.

Given the challenges that arise from analyzing historical documentation, scholars in the field of historical linguistics have recently made greater use of digital corpora. In the present diachronic study of Arabisms, for example, we will quantitatively analyze data obtained through searches of large corpora such as the *Corpus Diacrónico del Español* (CORDE), the *Corpus de Referencia del Español Actual* (CREA), the *Corpus del Español del Siglo XXI* (CORPES XXI) and the *Corpus del Español.* While these corpora offer little to no information on the factors of gender or age, they will provide valuable information regarding chronology, dialect, and register.

Regarding the reliability of documentation, there is always the question of whether it is ever truly possible to determine the date of incorporation or obsolescence of a lexical item. Not only is it possible for there to be discrepancies in the date of incorporation or removal of a particular lexical item depending on the corpora we use, it is likely in many cases that the item in question was used for years in speech before its written documentation. In addition, there is also the possibility that a given lexical item was not recorded in any dictionary or that its usage was not universal at a given point in time. For purposes of dating (and etymology), we will rely on the documentation of the CORDE due to its accessibility, size, and currency, as well as critical dictionaries such as Corominas and Pascual (1980-1991) and Corriente (2008). When we do not refer to a source explicitly, it can be assumed that dating of vocabulary comes from the CORDE.

Further methodological questions include the need to address the following concerns of any lexical study of languages in contact: (1) means of transmission, (2) motivations for borrowing, (3) degree of integration, (4) type of transmission, and (5) controversial etymologies. By means of transmission, we refer to the broad differences that result from borrowing from either speech or writing (Dworkin 2012: 8-14). What this generally means for Arabisms is a higher degree of orthographic variability for (certain) orally transmitted lexical items (e.g., *dragomán*, *drogmán*, *truchimán*, *trujamán*, *trujimán* 'interpreter') compared to those transmitted through the written register (e.g., *nadir* 'nadir').

A second distinction, between necessary (or cultural) and luxury (or core) borrowing, must also be made. As their names suggest, necessary (cultural) borrowing occurs when speakers of a given language do not have a word for a particular object or concept whereas luxury (core) borrowing occurs in spite of the existence of a word with a similar meaning (Goddard 1969, Dworkin 2012). No matter how close their meaning, however, a luxury borrowing is rarely perfectly synonymous with an inherited equivalent form (Dworkin 2012: 9). It is for this reason that it is important to consider any denotative or connotative differences between inherited and borrowed forms, any semantic changes they experience over time, and any social factors responsible for the borrowing in the first place.

A third distinction deals with the degree of integration of a loanword. Unintegrated borrowings are those recognized as foreign and maintain their original orthography, phonology, or morphology. Integrated borrowings are those that have been adapted to the norms of the recipient language. Upon full integration, a loanword undergoes derivational processes and

semantic changes and, thus, is treated as any other element of the inherited lexicon (Dworkin 2012: 10). Due to the presence of word initial *al-* in many Arabisms, speakers have been aware of the contribution of Arabic to the Spanish language. Of course, this oversimplification is quite misleading; there are words of Arabic origin without word-initial *al-* (e.g., *taza*), as well as patrimonial words with word-initial *al-* (e.g., *alegoría*). However, this misguided interpretation has had real consequences on words of Arabic origin, with language purists condemning certain words based on their form.[6]

Another important distinction we recognize is between immediate (direct) and remote (distant) borrowings, which leads us to one of the principal concerns of studying Arabisms, the definition of *Arabism* (Giménez-Eguíbar 2011: 37), given that Arabic has entered into Spanish in many different ways. For our purposes, the definition of *Arabism* as it pertains to Spanish is simple: An Arabism is any word having passed directly from Arabic or Hispano-Arabic into Spanish. Naturally, this means that the definition applies to any word whose original source was Arabic or Hispano-Arabic (e.g., Sp. *jinete* < Hispano-Ar. *zanáti*). We also include in this definition any word originating in another language but that entered into Spanish via Arabic or Hispano-Arabic (e.g., Sp. *almoraduj* < Hispano-Ar. *almarda[d]dúš* < Classical Ar. *marzanǧūš* < Gr. *ἀμάρακος*).

Finally, we must acknowledge that there is often disagreement regarding the etymology of borrowings. First, there is debate in determining the source language (e.g., *matar* from Latin or Arabic). In other cases, scholars may agree on the source language but disagree on the precise

[6] For further discussion of characteristically Arabic features, see section 2.3.4.3.3.

etymon (especially in the case of Arabisms) (Dworkin 2012: 14). In the present study, we will not provide an exhaustive explanation when presented with controversial borrowings; instead, we will present relevant information from lexical studies and dictionaries as appropriate.

2.3.2 Theoretical considerations in research on lexical variation in general and with regard to Arabisms

When performing an analysis of vocabulary, an important consideration is the difference between types of lexical variation. In Escoriza Morera (2002), for example, the author makes the following distinction:

(1) Variants of content; consider how *hasta*$_1$ ('until') and *hasta*$_2$ ('even, including') are variants of the same variable of expression (i.e., *hasta*).
(2) Variants of expression; consider how *olivo* and *aceituno* are variants of the same variable of content (i.e., 'olive (tree)').

The above distinction is what are traditionally known in structural linguistics as the studies of semasiology and onomasiology. Geeraerts, Grondelaers and Bakema explain that semasiology analyzes a word and the various meanings it takes on whereas onomasiology analyzes a concept and the various forms it takes (1994: 3).

Whereas most studies focus on onomasiological variation, there have been several important studies in recent years that address semasiological variation. One such work is that of Geeraerts, Grondelaers and Bakema (1994), which addresses lexical variation in items of clothing in Dutch.

While most semasiological studies rely on qualitative differences between variants, these researchers addressed whether different variants had the same structural weight (e.g., degree of representativity). In addition to discovering that lexical variants did not have the same structural weight, they found that relevant factors were both cognitive (e.g., prototypicality, or the degree to which a referent represents a given category) and contextual (e.g., diaphasic, geographical, and social restrictions) in nature. Since traditional semasiology has tended to ignore contextual factors and to focus solely on cognitive factors, future semasiological studies would do well to address them.

Grondelaers and Geeraerts have also made strides in the study of onomasiological variation. In Grondelaers and Geeraerts (2003), the authors make a distinction between conceptual and formal variation. With conceptual variation, different words may be used to express a given meaning, without their meaning being necessary identical. Consider, for example, that in Peninsular Spanish *alubia* 'bean' (< Ar. *allúbya*) is a word given to the edible seed of various legumes but that other alternatives (e.g., *haba* 'broad bean, fava bean' < Lat. FABA) for a given context are available. With this in mind, we are not suggesting that these terms are identical (i.e., true synonyms); obviously there are relationships of hypernymy and hyponymy between related, but not identical, words such as these.

In discussing formal variation, Grondelaers and Geeraerts (2003) explain that there are various ways to name a single referent. Naturally, this is a matter that has been highly disputed by prescriptive grammarians, who have been preoccupied with the economy of language and who have opposed redundancies in language. This preoccupation has led some to believe that

true synonymy does not exist, which would suggest that languages may contain words with similar, but not identical, meanings. However, it is clear that all languages do, in fact, maintain a degree of redundancy, as in Eng. *cantaloupe* and *muskmelon* or Sp. *alhucema* (< Hispano-Ar. *alḫuzáma* < Classical Ar. *ḫuzāmà*), *espliego* (< Lat. SPICŬLUM), and *lavanda* (< Fr. *lavande* or It. *lavanda*).

The kind of synonymy with which a number of sociolinguistics are primarily concerned is of the latter type, that is, formal variation, where words such as *alhucema*, *espliego*, and *lavanda* may be considered true synonyms of one another and, therefore, variants of a single lexical variable. Although it may be argued that the distribution of these words is not identical (which would make them ineligible as variants of the same variable), some researchers claim that their differences of distribution may be neutralized in usage (Escoriza Morera 2002). A brief perusal of any Spanish dictionary will easily demonstrate the equivalence of *alhucema*, *espliego*, and *lavanda*, with each word of the set having at least one meaning in common. Synonymy is also possible where only a secondary meaning of a given word is synonymous with another word. Such is the case of *alfóncigo* (< Hispano-Ar. *alfústaq*) and *pistacho* (< Fr. *pistache* or It. *pistacchio*), where the primary meaning of *alfóncigo* is ‘pistachio (tree),’ while the primary meaning of *pistacho* is ‘pistachio (seed),’ a secondary meaning of *alfóncigo*. As further support of the synonymy of words that are not identical, authors such as López Morales (2004) have stated that context may neutralize differences of words that one might consider very different but given the circumstances may act as synonyms (i.e., functional synonyms). In his work, the author recalls a study in which a group of students were instructed to vary the vocabulary they used in a piece of writing about their vacations. The experiment found that *playa* was regularly replaced

with *costa* and *mar*, two words considered to be conceptually different, supporting the idea that words with differing denotative meanings may, in fact, behave as synonyms of one another.

Another concern for skeptics of lexical variation is whether certain words, with significant differences in connotation, may be considered variants of the same lexical variable. Consider, for example, that death is generally taboo in many cultures and that speakers tend to use euphemisms when referring to it. For this reason, there are other words to express '(to) die' in addition to the most typical option, *morir*. In the speech of Castellón, the most frequent euphemistic alternative to *morir* is *faltar*, followed by *fallecer*. Naturally, the only way to prove that *fallecer*, *faltar*, and *morir* are variants of the same variable is to gather empirical data showing that they all appear in different styles and with the same denotative quality. It is even more convincing to find this sort of variation in the same speaker (Blas Arroyo 2009b: 196).

Other examples of word pairs having related meanings with different connotations are numerous, especially where Arabic-Romance lexical groups are concerned. As a result of negative attitudes toward words of Arabic origin, many words identified as such by native Spanish speakers developed negative connotations that their Romance counterparts never had, a fact that is especially true of Arabisms referring to Arabic speakers, culture, or customs. Sometimes this depreciation in meaning is obvious (e.g., *árabe* 'Arabic' vs *algarabía* 'Arabic; unintelligible language'), but in other cases the differences are subtler. For example, in the case of *alarife* (< Ar. *al'arif*) and *arquitecto* (< Lat. ARCHITECTUS), we have a lexical item, *alarife*, initially meaning 'architect, contractor,' but which later acquired meanings 'construction worker; mason; carpenter,' where there is a distinction in rank between the initial meanings of the word

and subsequent ones (Maíllo Salgado 1998). However, despite the potential depreciative connotation of *alarife*, not all speakers used it in that way,[7] which supports the idea that, in the proper context, two words with different connotations (e.g., *alarife* and *arquitecto*) may function as synonyms.

2.3.3 The role of linguistic factors in lexical variation

Although linguistic, or structural, factors are considerably less frequently analyzed in the study of lexical variation, some researchers have addressed the influence of surrounding context to determine the selection of variants. In Nadasdi (2005), for example, the author analyzed lexical variation in the French of Ontario, finding that, in conversation, the variant chosen by one speaker greatly influences the variant chosen by another speaker in his or her response. This conditioning of variants in context would also apply to written language, and in theory would be helpful in our understanding of the diffusion of a lexical variant. However, given the nature of the data we will analyze (i.e., the paucity of written exchanges in the documentation), we will not consider this feature in detail.

Another, and perhaps the most significant, linguistic factor is any that deals with semantic differences of lexical variants. In some cases, semantic differences affect the choice of one variant over another, as in the case of *aquí* and *acá*. In Caracas speech, Sedano (1994) demonstrates that *aquí* is more precise a locative adverb than *acá*, a fact confirmed in many instructional texts of Spanish, including Butt and Benjamin (2004: 444). Also, as stated above, Grondelaers and Geeraerts (2003) suggested that the selection or preference of a variant over

[7] The *Diccionario del español medieval* defines *alarife* as ‘architect, contractor’ but also ‘bricklayer, construction worker.’

another is largely conditioned by the prototypicality of a given semantic category; that is, a referent is more often expressed by a word that is more prototypical (i.e., semantically broader) in its particular lexical category. As a result, semantically broader terms tend to be more resistant to loss than related ones that are more semantically restricted.

As far as prototypicality in Arabic-Romance lexical groups is concerned, one example of interest is the semantic category of 'pillow,' in which *almohada* (< Hispano-Ar. *almuḫádda*) is the most prototypical. In this category, we also have derivatives *almohadón*, *almohadilla*, along with Latin forms *acerico*, *cojín*, and *cojinete*. As the most prototypical member[8] of this semantic category, it is not surprising that *almohada* is also the most dominant (i.e., frequent). What is interesting, however, is that Latin-based OSp. *faceruelo* was once equally prototypical in its meaning of 'pillow' but that it is no longer used in contemporary speech. One possible explanation for the disappearance of *faceruelo* in the face of *almohada* is a preference for certain goods made by Arabic speakers (Dworkin 2012: 96). While this claim is as yet unconfirmed, what is clear is that differences in prototypicality cannot fully explain the fate of *faceruelo* and *almohada*.

While linguistic factors are often informative, they alone cannot adequately explain the coexistence of competing Arabic-Romance lexical items. It is for this reason that we subscribe to the approach of informational maximalism, making use of any available resources in order to provide the best possible account of the preservation (and introduction) of the Arabisms

[8] The prototypicality of *almohada* is confirmed in the definitions of *almohadón*, *almohadilla*, *acerico*, *cojín*, and *cojinete* themselves, which support the following hierarchy of prototypicality: (1) *almohada*, (2) *almohadilla* / *almohadón*, (3) *acerico* / *cojín* / *cojinete*.

considered in our study. In the following sections, we will see how extralinguistic factors inform our analysis as well. Later, in the following chapters, we will see how an examination of both historical and modern texts inform our understanding of the lexical groups in question.

2.3.4 The role of extralinguistic factors in lexical variation

While structural factors are no doubt relevant for the variationist sociolinguistic approach to lexical variation, it is clear that lexical variation cannot be fully understood without the consideration of extralinguistic factors, by which we mean factors related to diachronic, dialectal, diaphasic, and social differences. Below we will consider extralinguistic factors in more detail and explain how they are relevant to the study of lexical variation in particular. Specific factors we will address are dialect (section 2.3.4.1), register and style (section 2.3.4.2), and social variables (section 2.3.4.3), which consists of subsections on gender (section 2.3.4.3.1), age (section 2.3.4.3.2), and language contact (section 2.3.4.3.3).

2.3.4.1 Dialect

Before the appearance of modern sociolinguistics, fields such as dialectology were already invested in the study of lexical variation. While dialectology was concerned almost exclusively with regional differences, research in this discipline has greatly contributed to other fields of linguistics. Alvar, in his contributions to Spanish dialectology, for example, employed sociolinguistic strategies such as choosing participants of both genders and choosing participants in such a way that they were representative of the population under examination (1973). In recent years, advances in technology have greatly facilitated the creation of linguistic atlases and the ability to gather statistical information using enormous amounts of data on lexical variation

where, in the early years of dialectology, scholars might have seen a disordered array of variability.

One such instance of the ability to access statistical information on a large scale is demonstrated in Moreno Fernández (1999). The study, which uses data contained in the *ALEA*, analyzed semantic category 'cowbell' and determined the geographic distribution of equivalent lexical variants of *cencerro* (e.g., *cencerra*, *piquete*). His quantitative analysis of this lexical category found that *cencerro* and its variants did not contain the same denotative quality throughout the Andalusian community, and it also established how groups of locales were organized into dialects and where dialectal boundaries occurred.

Although dialectal variation of the contemporary Spanish lexicon has recently had its share of attention in the field of variationist sociolinguistics, this type of variation has also had considerable influence in the history of the language. Consider, for example, how dialectal differences played an important role in the preservation of several words of Arabic origin despite the existence of perfectly equivalent Romance counterparts. Specifically, complementary distribution explains several cases of concurrent Arabic-Romance lexical variants. In many such lexical pairs, the Arabism is preserved in a specific region while the Romance term is part of the standard and is, therefore, found in general use throughout the rest of Spanish-speaking world. This fact is true of the aforementioned *alfóncigo* and *pistacho*, where *alfóncigo* is documented exclusively in Spain and *pistacho* is found elsewhere. This is not to say that *alfóncigo* is by any means popular even in the Iberian Peninsula, where *pistacho* is still considerably more common in both speech and writing. However, this fact does support the idea that when synonymous

Arabisms and Romance terms were not in intense competition with one another, the Arabism was able to endure, if only regionally.

2.3.4.2 Register and style

Diaphasic variation, or variation of speech itself, is often associated with differences brought about through changes of register or differences of style. As with dialectal variation, diaphasic variation was a research interest of scholars before the rise of the field of sociolinguistics. Ullmann (1962, 1964), for example, was interested in stylistic variation in sets of English words, such as *begin* and *start*. Despite earlier efforts to explain the roles of register and style in lexical selection, however, sociolinguistics offers greater insight on diaphasic variation, some of which we will see below.

While the effects of style and register on lexical choice are evident, defining their limits poses a challenge to research that has never been resolved in the field. One such attempt to address this difficulty is the explanation of Borrego Nieto (1994), that there are no finite distinctions between styles or registers but rather a continuum of formality along which the different levels of language, especially the lexicon, appear. In another study, Borrego Nieto (1981) analyzed the effects of the introduction of standard vocabulary into the local, nonstandard lexicon of the speech of a rural community in Zamora. Through the collection of data on frequencies and judgements on the words selected for the study, the author was able to establish that the selected words were located on different points along a continuum of formality, from lesser to greater formality. Furthermore, the study supports the idea that the use of one word at a

given point of formality does not necessarily mean that it will be accompanied by equivalents at the same degree of formality from another lexical category.

Escoriza Morera (2002) employed a different strategy to address the difficulties of diaphasic variation. In his study of the speech of Cádiz, the author asked that participants identify the lexical variant that they would employ under certain circumstances, from lesser to greater formality (e.g., joke, personal letter, announcement, complaint), and with whom (e.g., friends, partner, children, doctor, strangers). The author found that certain lexical variants were clearly associated with a high level of formality (e.g., *iniciar* 'begin') while others were clearly linked to a lower level of formality (e.g., *pasta* 'money'). Other lexical items were found to be neutral with regard to formality (e.g., *malo* 'bad'); that is, some words did not appear to be determined by the factors of register or style at all. Overall, Escoriza Morera (2002) found that diaphasic factors were the most relevant in his selection of vocabulary, even more so than social factors.

Building upon ideas discussed in Escoriza Morera (2002), Fernández Smith and Escoriza Morera (2004) investigated what the authors termed contextual adaptation (*adecuación*), that is, how speakers adjust their speech (and vocabulary) to fit a given context. The analysis found that the appropriate use of variants associated with greater (e.g., *ebrio*) and lesser (e.g., *borracho*) formality was significantly higher than the lack of adaptation (*inadecuación*). In other words, speakers are acutely aware of any changes in register and, accordingly, favor adaptation of lexical choice.

As far as Arabic-Romance lexical groups are concerned, one example that reflects well the differentiation of variants according to register is the semantic category of 'headache.' In this word group, we may characterize the available words (i.e., *cefalea, dolor de cabeza, hemicránea, jaqueca, migraña*) as either more or less formal along the continuum of formality. As words characteristic of popular usage, Latin-based compound *dolor de cabeza* 'headache' and Arabic-based *jaqueca* 'migraine' are found across a variety of different registers but are, more generally, associated with a lower level of formality. On the other hand, terms *cefalea* 'headache' and *hemicránea* 'migraine' (and to a lesser extent, *migraña* 'migraine'), which belong to the linguistic domain of medicine, are generally associated with a higher level of formality. While there is no retained Arabism with general meaning 'headache,' *jaqueca* (< Hispano-Ar. *šaqíqa*) successfully exists alongside Latin *migraña* and *hemicránea*, a case in which we see a popular Arabism appear alongside elevated Latinisms, a phenomenon paralleled by other lexical pairs (e.g., *alarife, arquitecto* or *albéitar, veterinario*). In other words, despite competition from *hemicránea* and *migraña, jaqueca* has thrived, a fact perhaps best explained by a difference of register.[9] We may attribute the register and preservation of *jaqueca* to its popularity as a subject of conversation in colloquial language. Its popularity as such is reflected in the fact that *jaqueca*, a term in the semantic category of health, is far better represented in the categories of narrative prose and fiction than in the categories of scientific prose and health, both diachronically and synchronically.

2.3.4.3 Social variables

[9] For information about the different types of transmission (i.e., direct, indirect) leading to the incorporation of Arabisms into Spanish, see section 2.3.4.3.3, on the role of language contact in lexical variation.

While the findings of Escoriza Morera (2002) showed that social factors were of secondary importance to stylistic factors, we cannot discount the influence that both exercise on one another. The field of sociolinguistics has verified time and again that, in the same communicative contexts, more formal variants are more often used by those of higher social classes and also by the middle classes and women, who are more influenced by linguistic norms and the prestige associated with them (Blas Arroyo 2012).

One example of research demonstrating the association between stylistic and social factors is that on taboos and their euphemistic or dysphemistic equivalents. López Morales (2005), who analyzed such terms in the speech of a Puerto Rican population unsurprisingly confirmed that the appearance of euphemisms increased as did the formality of the communicative situation. Less obvious, however, is the finding that social groups were not affected equally by stylistic factors. More specifically, he found that learned and technical words were most common among the middle class, followed by the upper class, and finally the lower class, which generally favored taboo equivalents. López Morales (2005) also found that other factors, such as age and gender, had a significant impact on the lexical variants selected.

Regarding the relevance of age and gender of informants, Pérez Vidal and Díaz Peralta (1996) analyzed the speech of a community in the Canary Islands and found significant differences in lexical selection between men and women, on one hand, and speakers of different ages, on the other. In their study, the use of euphemisms and dysphemisms among men is simply linear: younger men were shown to employ euphemisms and dysphemisms most often, then middle-aged men, and finally older men, who employed them least. The group of women was

statistically quite different in that the use of euphemisms is linear, by which we mean that the older the woman, the more frequent she would use a euphemism, but their use of dysphemisms is different: middle-aged women used dysphemisms the least, in contrast to both younger and older women, who used them significantly more often. Indeed, middle-aged women (but also middle-aged men, in general) experience greater sociolinguistic pressures (due to work and family roles) than either of the other two age groups (Holmes 1992: 186).

In the following three sections, we will discuss some of the social factors relevant in lexical variation research and how they may be applied to words of Arabic origin. Section 2.3.4.3.1 will address the role of gender, section 2.3.4.3.2 will discuss the factor of age, and section 2.3.4.3.3 will consider the effects of language contact, which are especially relevant to our study of words of Arabisms.

2.3.4.3.1 Gender

In the field of sociolinguistics, there has been some debate regarding the differences in language use in men and women, including the terminology of the variable itself (Blas Arroyo 2012). Without entering into too much detail, suffice it to say that an examination of the sociolinguistic literature of Spanish and English seems to suggest that the term *sexo* 'sex' is the preferred term in Spanish while in English it is *gender*.[10] Terminology aside, the linguistic differentiation between men and women parallels the other levels of linguistic analysis. In other

[10] We recognize that sex is biologically determined and that gender is socially constructed (Cheshire 2002). For this reason, gender better reflects differences in language use than sex. For our purposes, however, we will apply *gender* to both concepts since it is only in the past few decades that (some) sociolinguistic researchers have begun to acknowledge the distinction.

words, all other factors being equal, men typically employ more nonstandard variants than women, who tend to favor standard terminology (Trudgill 1983: 161).

Some studies have observed that certain vocabulary tends to be favored differently by men and women (García González and Coronado González 1998). Such is the case of words like *mono* 'cute,' which is a variant of *bello* / *bonito* / *lindo* that is used significantly more often by women (58%) than men (42%) in Cádiz, who favor the other variants (Escoriza Morera 2002: 269). Others claim that differences in the speech of men and women may be due to differences in discursive content rather than a difference of preference for words of a certain connotation. Lappalainen (2004), for instance, observed how a female Finnish speaker employed conversational markers much more often than her male interlocutor in all communicative contexts, which the author explained may also have been due to the especially talkative nature of the female speaker, rather than a pattern of linguistic variation due to gender. On the other hand, the preference for certain characteristic terminology could also be explained by different societal expectations exerted on the speech of men and women (Eckert 2000).

While gender is often found to be a significant factor in synchronic analyses of the lexicon, it is much more challenging to analyze the role of gender in the lexicon from a diachronic perspective. The reason for this difficulty is that some of the largest available corpora for historical linguistics, CORDE and the *Corpus del Español*, are largely lacking in author information, and in cases where the author is anonymous there is no such information whatsoever. A further challenge is that the vast majority of the authors of the works in these corpora are men, which distorts results in a diachronic analysis of this factor. Given the

aforementioned limitations, however, whenever we find data on gender, we will include it as appropriate, and any information obtained will be evaluated carefully to determine its significance. In doing so, we are reminded again of the approach of informational maximalism, by which we will use whatever useful information we find to help us better understand the preservation of the Arabic-based lexicon in Spanish. This strategy, then, includes examining historical and contemporary documentation, as well as additional contemporary lexical studies to provide evidence in support of our claims.

If we consider the Arabic-Romance lexical group representative of semantic category 'insane, crazy,' we find an Arabism, *loco*[11] that coexists in contemporary Spanish with a large number of higher register words of Latin origin (e.g., *alienado*, *delirante*, *demente*, *enajenado*, *insano*, *lunático*, *vesánico* but also hybrid derivative *enloquecido*) and a group of lower register words of diverse origins (e.g., *atreguado* < *atreguar* < *tregua* < Goth. *trĭggwa* 'treaty,' *chalado* < *chalar* < Caló *chalar* 'go,' *chiflado* < *chiflar* < Fr. *siffler* 'whistle,' *majara* / *majareta* < Hispano-Ar. *maḥrúm* 'wretched,' but also *grillado* < *grillarse* < *grillo* < Lat. *GALLELLUS 'sprout'). Although *loco* is an Arabism, it has more in common, stylistically, with the Latin-based words than the more colloquial second group. Colloquial meanings aside,[12] given the tendency for women to use more formal variants and for men to use more informal ones, women are more likely than men to use *loco* and other high register variants from this semantic category.

2.3.4.3.2 Age

[11] Corriente (2008) posits a Hispano-Arabic origin: *loco* < Hispano-Ar. *láwqa* < Classical Ar. *lawqā'* 'stupid.'
[12] Although we acknowledge that *loco* has many expressive (i.e., nonliteral meanings), here we are referring only to its use as 'insane, crazy.'

Age is another highly studied social factor in lexical variation research, which is reasonable, because age is believed to be one of the most significant factors in speaker selection of vocabulary. Anyone who has reflected on the differences between their own speech and that of his or her grandparents (*couch*, *davenport* 'couch, sofa') has realized that different generations do not speak like one other, especially with regard to vocabulary. While much has been written on lexical variation among younger speakers, one way in which the study of age and variation could improve is to expand upon literature on the speech of older generations (Blas Arroyo 2012).

Many studies focused on the role of age deal with generational differences and how they may lead to linguistic innovations and change. Labov (1972), with his presentation of the apparent time hypothesis, claimed that different frequencies of usage of variants across generations may indicate a change in progress in a given speech community; while this is often true, it is not always the case. Different frequencies of usage may be only temporary since different age groups adopt and discard linguistic features at different phases of their lives, a phenomenon called age grading (Labov 1994). While the term often refers to linguistic features adopted by younger speakers and discarded upon their entering adulthood, age grading also includes late adoption, which is described below.

The counterpart to the synchronic studies of apparent time are those of real time, which analyze the speech of their participants over a given period of time and arguably have been better able to confirm or reject projections of changes in the lexicon. One such study analyzed differences among linguistic atlases of North America and found that, of the 75 lexical categories

studied, more than one-third showed significant differences of usage among generations; however, we must recognize that, in contrast to the other types of linguistic features, it is not uncommon for a speaker to adopt at an older age vocabulary that is popular at the time (Boberg 2004: 257). This phenomenon, called late adoption, is yet another outcome to consider when identifying a potential change in progress.

An Arabic-Romance lexical group for which age is a factor is that of semantic category 'lavender (plant),' with variants *alhucema*, *espliego*, and *lavanda*. Coexisting for centuries since the medieval period,[13] Arabism *alhucema* (< Hisp. Ar. *alḫuzáma*) and Latin-based *espliego* (< OSp. *espligo* < Lat. SPICŬLUM < SPICUM 'spike; ear') have historically dominated the Spanish language in their meaning of 'lavender,' but more recent borrowing *lavanda* (< Fr. *lavande* or It. *lavanda*) has become increasingly popular, particularly among younger generations.[14] CORPES XXI already confirms the near exclusivity of *lavanda* in contemporary Latin American Spanish, but even in Spain *lavanda* appears in more cases and documents than *espliego* and *alhucema* combined. With the predominance of *espliego* and *alhucema* in the language of earlier texts and the contemporary decline in their usage, we see a change in progress where these earlier terms are losing ground to *lavanda*.[15]

2.3.4.3.3 Language contact

[13] Variant *alhuzema* is first documented in 1381-1418, and variant *espligo* first appears in 1379-1425.
[14] In an informal survey we conducted of native speakers of Latin American and Peninsular Spanish between the ages of 25 and 35, all preferred lexical variant *lavanda*.
[15] Despite the existence of two well-established terms, Colón correctly predicts the success of the Romance borrowing at the expense of both the others: "...*alhucema* queda reducida al Sur mientras que *espliego* es la voz del castellano general (en espera de que ambas sean sepultadas por la moda tonta de *lavanda*...)" (1998: 227)

In recent years, there has been an interest in the somewhat more complicated effects of language contact on lexical variation in bilingual speech communities. Some of the more interesting research is that which focuses on communities whose actual usage of the two languages is noticeably different. One example of this is Ontario, a province that is officially English-French bilingual but where French is generally limited to use in academic settings and official documents in a large portion of its population.[16] Nadasdi and McKinnie, who analyzed lexical variation in the speech of students of French in Ontario immersion programs, discovered the following: restricted and non-restricted bilinguals alike employed fewer lexical variants than native French speakers; restricted bilinguals use far fewer informal variants and prefer formal variants that are not especially common in speech typical of French-dominant speakers (2003: 59). What this study demonstrates is that despite a high level of linguistic competence, students of non-native languages may not necessarily (and often do not) have a high level of sociolinguistic competence in the lexicon (Knaus and Nadasdi 2001: 303), a fact that undoubtedly applies to the other aspects of language.

Other studies on language contact have addressed how loanwords affect bilingual communities to different degrees. In some of these studies, researchers have observed how foreign loanwords in one language are used in a speech community where the foreign language is the dominant language of the speakers. Etxebarría (1985), for instance, analyzed Spanish in Bilbao and observed how words of Basque origin are considerably more common in the speech of native Basque speakers than in the speech of Spanish speakers who do not speak Basque

[16] The fact that speakers of bilingual communities do not necessarily use both languages equally underscores the need for researchers to distinguish between restricted and non-restricted bilinguals; in the case of Ontario, these terms refer to speakers with English as their dominant language (and French as a secondary language) and to speakers with equal competence in both languages, respectively.

natively. In other cases, the effect of borrowing on the speech community is less conspicuous. One such effect is covert interference, which is an increased frequency of usage of variants due to their resemblance to words of another language (Mougeon and Beniak 1991: 11) Explanations such as formal similarities or shared etymologies would drive, for example, the preference for French variants *juste* and *seulement* among native English and Spanish speakers, respectively (cf. Eng. *just*, Sp. *solamente*) (Mougeon and Rehner 2001: 411). Other studies of variation found in Canadian bilingual communities have demonstrated the degree that linguistic standardization has influenced the speech of bilinguals. In Aléong (1981), the author observed that, regardless of the pressures exerted by traditionalists in favor of variants of French origin, even in Quebec there is still a considerable number of words of English origin, particularly in the lexical category of sports, not to mention the more recent English borrowings in the semantic category of technology. While a significant portion of lexical variation research has focused on bilingual communities in English-French contact situations, studies on Spanish would be even more fruitful given the number of contact situations it maintains with other speech communities worldwide (Lipski 2010).

As in the above bilingual communities, an examination of language contact between Arabic and Romance is essential to understanding lexical variation in Spanish as it pertains to Arabic-Romance lexical groups. With the Muslim invasion of the Iberian Peninsula (711), speakers of Proto-Ibero-Romance (i.e., the result of contact between Late Latin and pre-Roman languages and, later, Germanic languages) first came into contact with speakers of Arabic (and

Berber).[17] As a result of this contact, Ibero-Romance gained a significant number of lexical items from Arabic, which we may classify broadly as either direct or indirect. Direct Arabisms are borrowings integrated into Ibero-Romance (1) with the arrival of Mozarabs[18] to the Christian kingdoms of northern Iberia and (2) from Christian conquerors coming into contact with Mudejars[19] and later Moriscos[20] during the Reconquest of the Peninsula. Indirect Arabisms are (1) words disseminated by individuals from various parts of the world who had had contact with Arabic and who introduced them into their own languages, including Ibero-Romance, and (2) words created through translations of Arabic texts. An additional, smaller group of indirect Arabisms consists of (1) words used among contemporary authors in exotic texts and (2) words generated during the Spanish presence in North Africa or those used in contemporary press (Corriente 2005: 185-188).

Since there are different possibilities for transmission type and chronology, we suspect that there is a difference in the retention of direct and indirect Arabisms. Given the written nature of many indirect Arabisms, it would appear that these indirect borrowings are more likely associated with an elevated register than direct borrowings, which are the result of contact through speech. Therefore, we should expect that indirect Arabisms are more resistant to loss than direct ones.

[17] In addition, before said contact, there existed a situation of diglossia in the written register for each of these groups: In the case of Iberian natives, the written language was Latin, whereas for Arabs and Berbers it was Classical Arabic (Corriente 2005: 186).
[18] Iberian Christians (once) living under Muslim rule.
[19] Muslims of Al-Andalus who remained in Iberia after Christian conquest but did not convert to Christianity.
[20] Iberian Muslims who converted to Christianity.

In addition to their manner of transmission (i.e., direct or indirect), another aspect to consider when studying lexical variation among Arabisms are the historically negative attitudes of Romance speakers regarding Arabic. More concretely, the more characteristically Arabic (or non-native) the word is, the more we expect Romance speakers to have suspected that the word came from Arabic. Especially in competition with a patrimonial or Romance-based word, speaker awareness of Arabic-like features should have had a significant impact on the survival of an Arabism. With this in mind, we will address any formal features that offer insight into the origin of our selected Arabisms, keeping in mind that identification as such using Arabic-like features is a subjective criterion. To support our claim, we will seek support in older prescriptive language texts (e.g., Gandavos, López Villalobos, etc.), which completely reject the use of Arabisms. Features identified as characteristically Arabic by educated grammarians of Romance may be broadly classified as phonological, morphological, syntactic, or semantic.

Phonological features suggesting an Arabic (or non-Romance) origin are word-final /í/, /x/, /m/, /t/ as in *nazarí, almoraduj, islam, cenit*. In the area of morphology, the most common feature is morpheme *-í,* which in Arabic was used as a demonym (e.g., *marroquí*) or an attributive (*jabalí* < Hispano-Ar. *ǧabalí* '(from the) mountain,' *sandía* < Hispano-Ar. **sandíyya* '(from) Sindh'). The most obvious characteristic of many Arabisms, however, is their incorporation of Arabic article *al-*, with unassimilated (e.g., *algodón*) or assimilated (e.g., *azar*) /l/. Naturally, in terms of semantics, many words that refer to Arabic or Muslim culture, language, or individuals would also be easily identified as Arabic in origin.

Of course, any one linguistic feature alone would not necessarily have been sufficient for Romance users to correctly identify the origin of a word,[21] but in cases such as *algarabía* (< Hispano-Ar. *al'arabíyya* 'Arabic, from Arabia') and *marroquí*, each example, respectively, contains identifiably Arabic features from three of the different linguistic levels, seriously removing doubt about their Semitic origin. In our analysis, we expect that it is more likely for an Arabism with fewer characteristically Arabic features to be maintained than an Arabism with more such features.

2.4 The study of preservation of Arabisms in Spanish

While lexical variation has been more of a focus in variationist sociolinguistics in recent years, it still is significantly less popular in the scientific literature than, for instance, phonological, morphological, or syntactic variation. Some of the reasons why lexical variation has not shared the same degree of popularity areF that, compared with the other levels of linguistic analysis, there are significant methodological challenges, such as access to an adequate amount of data on variables, and also theoretical difficulties, such as encountering variables with meanings that are equivalent.

As a way to compensate for one of the major methodological challenges of the study of lexical variation, the relatively low frequency of lexical items in natural speech, sociolinguists have employed questionnaires or interviews to elicit variants from participants. One disadvantage in employing these strategies, however, is the observer's paradox, especially in an unnatural communicative context. In order to avoid it, some researchers have chosen to analyze

[21] Consider, for example, OSp. *abezar* (MSp. *avezar*), wrongly accused of being an Arabism due to its presence of word-initial *a-* (Giménez-Eguíbar 2011: 7).

data from oral corpora. The use of the questionnaire, the interview, and oral corpora in conjunction with the fact that there are lexical variants with shared semantic meaning will continue to be essential in the future study of an area of sociolinguistics that deserves more attention than it has received thus far.

Now, while questionnaires, interviews, and data from oral corpora are useful in a synchronic analysis of contemporary speech, the same is not true for a diachronic study of an early speech form or of writing. In the absence of available speakers or oral data, researchers must seek alternatives. In recent years, historical linguists have turned to written data found in corpora such as CORDE, CREA, and the *Corpus del Español*. In order to conduct a diachronic analysis of the Spanish lexicon for the present study, we will primarily use data available in CORDE and CREA, in conjunction with different forms of documentation, namely, critical dictionaries and lexical studies.

As far as theoretical challenges are concerned, we must be certain that a proposed set of variants are truly members of the same variable (given the context) or, on the other hand, if there are certain linguistic or sociolinguistic restrictions that prevent their synonymy. In any case, studies such as those addressed above suggest that synonymy is more common than one might expect without close examination. It is important to recall that variants do not have to be true synonyms of one another to carry the same denotative value; even partial synonymy in conjunction with neutralization of semantic differences in context allows apparently different words such as *costa* ‘coast,’ *mar* ‘sea,’ and *playa* ‘beach’ to function as synonyms (i.e., functional synonymy).

Regarding the matter of synonymy in vocabulary, all lexical items selected for the present study have at least one definition in common, as confirmed in dictionaries or lexical studies. In a lexical pair such as *escorpión* (< Lat. SCORPĬO, -ŌNIS) and *alacrán* (< Hispano-Ar. *al'aqráb*), for example, both share meaning 'scorpion (arachnid)' and, accordingly, are synonymous when found in contexts referring to scorpions. Obviously, both terms have numerous distinct definitions, but this fact does not prevent them from functioning as synonyms when used in the right contexts. Since identifying these contexts is necessary for determining synonymy within a set of vocabulary, we must define the envelope of variation[22] for each of the lexical groups analyzed in the following chapters. Given the complexity of studying a diverse group of lexical variables over a period of centuries, we will describe the envelope of variation of our Arabic-Romance lexical groups on a case-by-case basis.

Despite the challenges of studying lexical variation, much of what we have learned about variation at other levels of linguistic analysis may also apply to vocabulary and to the study of Arabic-Romance lexical groups in particular. Namely, many studies analyzing lexical variation have found evidence mirroring the results that studies on other types of variation have reported, such as the accommodation of vocabulary based on different styles or the use of nonstandard terminology by the groups of speakers who are least affected by prestige norms (i.e., lower social classes, men, younger speakers) or, on the other hand, the preference for standard vocabulary by those who most often subscribe to them (i.e., upper social classes, women, older speakers).

[22] Defining the envelope of variation, or circumscribing the variable context, is the determination of contexts in which a given linguistic variable occurs. As well as identifying the contexts that variants share, the researcher must exclude contexts in which items do not function as variants of the same variable (Tagliamonte 2006: 86).

With the knowledge of previous variationist studies in mind, we will analyze relevant linguistic and sociolinguistic factors to determine the extent to which they work together to explain the preservation (and coexistence) of words of Arabic origin in the Spanish language. Out of the linguistic factors, we believe the most relevant in the maintenance of Arabisms is semantic differences. Of greater importance are sociolinguistic factors, of which we will take into account dialect, register / style, and social factors. Specifically, we expect to find the following in our analysis: (a) that regionalism is favorable to the maintenance of Arabisms, (b) that an Arabism is favored if it appears on the opposite side of the spectrum of formality from its Romance or Latinate counterpart(s), and (c) that neutral (or favorable) attitudes toward the Arabism in question increase its likelihood of survival. Given that the Arabisms we analyze come from historical works of the written register, composed almost exclusively by male authors, we do not believe gender is a significant factor for maintenance of Arabisms. Addressing the factor of age will not be definitive and is only possible in a broad sense in that we may only speak of trends over a large period of time since biographical information on such texts is largely unavailable; however, we will discuss any relevant trends as appropriate. Once more, in line with the idea of informational maximalism, we will incorporate in our analysis any information pertinent to the preservation of competing Arabic-Romance lexical items. In our case, this will entail reviewing both historical and contemporary texts for clues on word usage, restrictions, and change, in addition to evaluating how different varieties have employed the lexical items over time, if at all.

While researchers have made an effort to address some matters concerning words of Arabic origin, we intend to address the absence in the literature regarding the resilience of Arabisms that coexist alongside equivalent lexical items of Latin or Romance origin. The objective of the present study, then, is to address this resilience employing both historical and contemporary sociolinguistic methodology. In doing so, we intend to explain the retention of a set of Arabisms through an examination of their sociohistorical context and relevant linguistic and sociolinguistic factors.

3 Case study: *aceite, olio* and *óleo*

In this chapter, we present our first of three Arabic-Romance lexical groups, namely, that of *aceite, olio,* and *óleo,* three functional synonyms that we have chosen to represent the broader domain of agriculture.[23] By analyzing a number of linguistic and sociolinguistic factors, as well as the historical context in which the olive and olive oil became (re)introduced into the Iberian Peninsula, we will explain which of these aspects are favorable, unfavorable, or neutral to the retention of the Arabism (i.e., *aceite*). We contend that the introduction of *aceite* corresponds to a perceived need to identify a new cultural reality for Castilian speakers and that retention of the term is supported by the semantic differences that it has maintained from its patrimonial counterparts, as well as generally neutral attitudes toward the Arabism. The introduction of the near synonym *óleo* forms part of the lexical transition of Castilian that occurred during the second half of the fourteenth century and the first half of the seventeenth century. In more general terms, our goal is to better understand which of these conditions allowed agricultural Arabisms to endure despite the loss of prestige of Arabic and the existence of synonymous patrimonial terms.

Below, we briefly introduce Arabisms in the domain of agriculture, before continuing on to section 3.1, where we describe the historical context of the olive and olive oil, including its degree of cultivation, its economic and commercial significance, and its social consequences in medieval Iberia. In section 3.2, we determine the impact that the linguistic factors of formal

[23] In later sections, it will become clear that these terms are not exclusively agricultural in nature. However, given that all three of the terms refer to products that are derived or cultivated for human use, the domain of agriculture is appropriate.

variation, related vocabulary, and semantic differences have had on the retention of *aceite*. In section 3.3, we address the sociolinguistic factors of dialect, register, and language contact to the same end. In our analysis of factors, we primarily use textual evidence from CORDE (for formal variation, semantic differences, and register), historical and contemporary dictionaries (for related vocabulary), and lexical studies (for dialect and language contact). The final section of this chapter, 3.4, will provide a summary of the results.

While only a few agricultural Arabisms are introduced into Hispano-Romance during the first wave of Arabic loanwords in the eighth to eleventh century,[24] many more are borrowed in a second wave in the twelfth and thirteenth centuries, an age marked by the major campaigns of the Reconquest. In direct contact with Arabic, Hispano-Romance incorporated of number of Arabisms associated with Arab practices and products, many of which represented various aspects of the countryside. These borrowings included words that identify food products (*acelga* 'chard,' *arroz* 'rice,' *azafrán* 'saffron,' *azúcar* 'sugar,' *bellota* 'acorn'), as well as other plants and flowers (*adelfa* 'oleander,' *arrayán* 'myrtle,' *nenúfar* 'water lily,' *jara* 'rockrose'), subcategories that are not represented in the first wave of Arabisms (García González 2007). Many more agricultural Arabisms (e.g., *albaricoque* 'apricot,' *limón* 'lemon,' *naranja* 'orange') are introduced in the following centuries, representing one of the most numerous and successful (i.e., preserved) categories of Arabic loanwords (Maíllo Salgado 1998).[25]

[24] Refer to García González (2007) for an explanation of the division of Arabisms documented during this time period (711-1300).

[25] This is only intended as a brief overview. For further reading regarding the Arabisms introduced during this period, we recommend Neuvonen (1941) (eighth through thirteenth century) and Maíllo Salgado (1998) (1300-1514). Corriente (2008) also discusses Arabisms in Ibero-Romance as a whole, from the invasion of the Iberian Peninsula to the present.

3.1 Agriculture and commerce in Iberia, especially concerning the olive

Approximately 5,000 to 6,000 years ago in the eastern Mediterranean,[26] the European olive (i.e., *Olea europaea*) was first cultivated with great success (Chandler 1950). As it became domesticated throughout much of the Mediterranean region, this evergreen tree became prized for its wood, and especially for its fruit and the oil extracted from it. As an important commodity in the ancient world, olive oil was in high demand (Tannahill 1988). In regions with favorable climates (e.g., Greece, Italy, Spain), the olive tree was lucrative despite the fact that the tree often required several generations to bear fruit. Once the olive matured, cultivators could expect the resilient tree to live and produce fruit for many hundreds of years, barring some natural disaster; the centuries-old olive trees in Spain and other areas of the Mediterranean are certainly a testament to this (Chandler 1950).

Domesticated between the eighth and seventh century B.C., the olive would eventually become one of the most important crops of the Roman empire, as the product formed a significant part of the Roman diet and was of high economic value.[27] Of the expansive territory covered by the empire, the greatest producer of olive oil for Ancient Rome was Hispania Baetica.[28] Evidence of the importance of the olive is found in the works of authors from ancient Rome, including Cato, Pliny, and Virgil (Vaquerizo Gil 2011). Although the olive was one of the most important components of the economy of Hispania Baetica in Antiquity, the decline of the

[26] Chandler (1950) posits an origin in the Middle East (e.g., Palestine, Syria).
[27] Vaquerizo Gil (2011) states that the price of olive oil was very high. For example, half a liter of *oleum flos* (high quality olive oil) was equivalent to what a teacher would charge a student per month or, alternatively, the daily wage of a skilled artisan (e.g., baker, blacksmith, carpenter).
[28] (Hispania) Baetica, an ancient Roman province, roughly corresponds to modern Andalusia.

Roman empire and subsequent Visigothic rule in Iberia signaled a decline in agriculture (Vaquerizo Gil 2011) and cultivation of the olive (Picornell Buendía and Melero Martínez 2013).

However, in the eighth century, the Moors invaded and established themselves in the southern portion of the Iberian Peninsula and made improvements in this regard, introducing new plants (e.g., cotton, eggplant, lemon, orange, rice, watermelon, and others) and irrigation systems that allowed agriculture to flourish once again (Trillo San José 2007: 106-110). With access to new agricultural practices, the olive became very successful in Al-Andalus, no doubt due to the fact that the Moorish territory coincided with the area of the Iberian Peninsula that was most conducive to olive cultivation (Martínez Enamorado 2007: 169). In other words, with its largely arid and warm climate, southern Spain was ideal for the cultivation of the olive and production of olive oil. On the other hand, the tree was virtually or entirely absent from the wetter and cooler areas of the country (i.e., the North), even though the olive was already widespread throughout much of the Iberian Peninsula by the Middle Ages (Dufourcq and Gautier-Dalché 1983, Casado Alonso and Ruiz 2019).

The uses of the olive in al-Andalus were many, with evidence suggesting that its inhabitants grew olives primarily for themselves for consumption (Trillo San José 2007: 110), although they were also exported elsewhere in the Mediterranean, such as Crete, Egypt, the Maghreb, and Yemen (Martínez Enamorado 2007: 172-175). As a result, the olive was of considerable importance, not only to the culture of al-Andalus, but to the economy of the territory as well. In addition to its consumption and exportation, the olive was used for illumination, hygiene, medicine, and many other purposes (Vaquerizo Gil 2011: 662-664).

3.1.1 The Crown of Castile until 1300

While contemporary olive and olive oil production far exceeds the needs of the Iberian Peninsula, early Castilian Spain was not always able to provide for those refugees who, in the ninth and tenth centuries, fled the South to live in the North. In the new environment, these refugees often did not have access to certain products to which they had been accustomed in the South. Among the most important of these was olive oil, which was not widely produced in Castile at the time (Dufourcq and Gautier-Dalché 1983: 40). In the absence of olive oil, individuals sought alternatives, which included oils extracted from other available plant-based materials, including nuts (e.g., almonds, walnuts), seeds (e.g., flax, sesame), and flowers (e.g., rose, violet), depending on the intended purpose of the oil (Tannahill 1988).

In the twelfth and thirteenth centuries, agriculture had become an essential component of the Iberian economy and succeeded, in large part, to the fertile Castilian land and the dedication and skill of the inhabitants who cultivated it. Throughout the various kingdoms of the Iberian Peninsula, olives, grapes, and cereal grains were among the most important crops (Casado Alonso and Ruiz 2019: 3-4). By the turn of the fourteenth century, fruit-bearing trees had become a significant source of the rural economy in regions north of the Cantabrian mountain range, but there olive trees were relatively scarce. For this reason, all along the northern coast, oil was derived primarily from alternatives such as nuts (e.g., chestnuts), while another significant source of oil was fish. South of the Douro Valley, in the highland regions of the Castilian meseta, olive trees were present but not especially common. In Andalusia and Murcia, however, olive

groves were a common sight, flourishing in the warm, dry South (Dufourcq and Gautier-Dalché 1983: 119, 153, 154, 160, 168, 177).

3.1.2 The Crown of Aragon in the thirteenth and fourteenth centuries

In the thirteenth and fourteenth centuries, cultivation of the olive tree was so widespread and successful in Aragon, Valencia, and Majorca that these regions not only met their own needs, but they also were able to export olive oil (Vaca Lorenzo 2016). Cultivation of the olive tree was substantial; even in the driest zones of Aragon lush orchards flourished along the Ebro (Burns 1967). As a result of its agricultural surplus, Aragon exported olive oil to Navarre and France and exported many other products to Castile and the Mediterranean and Atlantic coastal regions (e.g., wine, oil, fruits, rice, honey, bacon, metal goods, textiles). In turn, Aragon was able to import several different types of goods from the East, including cotton, medicines, spices, sugar, and textiles. Even during the chronic deficit of cereal grains in thirteenth-century Iberia, Majorca was able to produce large quantities of oil; since production far exceeded the needs of the island, surpluses of these products were heavily exported to other areas of the Peninsula and beyond (Dufourcq and Gautier-Dalché 1983: 181, 182, 185, 237).

By the fourteenth century, the network of commerce in Barcelona, Majorca, and Valencia was truly global, with trade occurring between them and modern-day Turkey, Greece, Syria, Egypt, Cyprus, Italy, Libya, Tunisia, France, Algeria, Morocco, Portugal, Belgium, and England. Within this expansive trade network, the Crown of Aragon exported and imported a number of diverse products, such as metallurgical products, coral, jewels, spices, fabric, linen, as well as wine, oil, saffron, rice, dried fruits, oranges, pomegranates, and wheat (Dufourcq and Gautier-

Dalché 1983: 245).

3.1.3 The Crown of Castile from 1350 to 1474[29]

The period between 1350 and 1474 was marked by an increasing interest in commerce on the part of the Castilian high nobility. Recognizing the financial benefit of involvement in foreign trade, the elite heavily invested in the success of the fishing and olive oil trades, which were lucrative for Castile at the time (Dufourcq and Gautier-Dalché 1983: 249, 255). While individual nobles were permitted to buy and sell goods with other territories, significant trade was performed on a much larger scale as well.

From the fourteenth century onward, northern coastal communities, which extended from Galicia to the Basque coast flourished. Although fishing was a principal economic activity there, the region also relied heavily on agriculture. The North had established connections with Western Europe (e.g., England, Flanders) for the importation of various goods (García de Cortázar 1982). Exports from the region included iron and wool, and products that it re-exported from the South included dried fruits, olive oil, spices, and wine. Another region of immense commercial importance located further south was Andalusia, a region rich in natural resources that regularly produced surpluses of important products (e.g., cereal grains, rice, peas, garbanzos, fruits, saffron, oil, wine) under favorable conditions (Dufourcq and Gautier-Dalché 1983: 256-258, 259). Of these products, olive oil was one of the most lucrative (Vicens Vives 1969).

By the fifteenth century, Castilian commerce extended from Galicia in the North to

[29] This time period begins with the death of Alfonso XI and ends with the outset of the reign of Ferdinand and Isabel.

Murcia in the South, centered on two poles: Burgos-Bilbao and Seville (Vaca Lorenzo 2016). The Basques, then considered an intermediate point between northern and southern regions, provided Andalusia with iron, wood, leather, and fish; in return, Andalusia provided the Basques with oil, wine, citrus, almonds, rice, and wheat. Despite their shared connection in Bilbao, economic systems in the North and South were very different from one another. On the one hand, Seville was primarily concerned with trade in Africa, the Mediterranean, and the Atlantic; on the other hand, Burgos and Bilbao were primarily concerned with trade in England, France, and Flanders. Seville (alongside other southern ports) was a cosmopolitan market of intense financial activity. In the North, there was no such place, and foreigners were generally scarce (Dufourcq and Gautier-Dalché 1983: 262).

Equally as important to the economic stability of Castile was internal trade. During the second half of the fourteenth century and the entire fifteenth century, trade among the interior of Castile had three main functions: (1) supplying the regions that temporarily or constantly were in need of indispensable items (e.g., cereal grains, wine, oil), (2) distributing raw materials and the products manufactured from them (e.g., fabrics from Castilian textile centers), and (3) redistributing imported goods. In Castile, agricultural production was usually able to sustain its growing population, allowing for surpluses of several products (e.g., fruits, oil, wine, and cereal grains) to be exported. Although polyculture was common throughout the territory, certain regions specialized in different kinds of crops (Casado Alonso and Ruiz 2019: 4). For instance, vineyards thrived in many regions, especially in La Rioja, La Mancha, Cordoba, and Jerez; for the olive tree, the most productive region was Andalusia. Such specialization of agriculture occurred relatively early, as made evident by an example of a document from the fourteenth

century concerning an area surrounding Seville where olive groves occupied more area than wheat, barley, and vineyards combined (Dufourcq and Gautier-Dalché 1983: 263, 266, 267). In Andalusia, olive trees were the base of the economy and at times were the only crops that produced a significant profit (Vicens Vives 1969).

3.1.4 Overview of agriculture and commerce in medieval Iberia

Throughout this section (3.1), we have discussed the origin of the olive and olive oil, the extent of olive cultivation and olive oil production, and their economic importance in the Crowns of Castile and Aragon during the Middle Ages. For centuries, olive oil production was limited to the southernmost regions of the Iberian Peninsula, but once olive oil was introduced into Castile (and Aragon) in the early thirteenth century, the territory experienced an unprecedented economic boom led by olive oil production (Lodares 1992: 1147-1148). With the immediate and substantial economic benefit that olive oil provided, the term (*aceite*) that meant 'olive oil' spread as quickly as the new material reality itself. As *aceite* represented what would soon be considered an essential product for Iberian Christians, the Arabism, popularized and spread through social networks like the marketplace, cemented itself in the lexicon. These circumstances, that is, rapid, widespread production of olive oil and use of the corresponding Arabism, as well as the prior need of a word to fill what was considered a terminological void (i.e., a necessary, or cultural, borrowing), are highly favorable to retention.[30] In the following sections, we will explore other favorable factors, as well as those that are unfavorable, irrelevant, or inconclusive in the maintenance of this Arabism.

[30] For an explanation of what constitutes a necessary (or cultural) borrowing, see section 2.3.1. Further details of the transmission of this Arabism are found in the sections concerning register (3.3.2) and especially language contact (3.3.5). We discuss how *aceite* qualifies as a necessary borrowing in the section regarding semantic differences (3.2.3).

3.2 Linguistic factors

Having examined the historical context in which the olive and olive oil existed (section 3.1), we now turn our attention to the linguistic factors involved in the preservation of *aceite* despite the presence of an equivalent Romance counterpart. We will begin our discussion addressing formal variation (3.2.1), before addressing related vocabulary (3.2.2), and finally semantic differences (3.2.3). Generally, limited formal variation, abundant related vocabulary, and notable semantic differences are all favorable to the preservation of an Arabism. As we will see below, not all of these linguistic factors are equally favorable to the retention of *aceite*, nor is it necessary that all factors be favorable for an Arabism to endure.

3.2.1 Formal variation

One of the most conspicuous features of historical texts is their significant formal variation, within individual authors, translators, or scribes. Studies suggest that considerable variation in form is unfavorable, capable of displacing even patrimonial terms (Malkiel 1976, Dworkin 1989, Cano Aguilar 1993).[31] If formal variation is unfavorable to inherited terminology, then Arabisms are even more vulnerable. In other words, the greater formal variation that *aceite* presents, the more susceptible it is to elimination. Let us first consider four of the earliest orthographic variants of *aceite* below.

[31] For example, Malkiel (1976) contends that the elimination of reflexes of Lat. FIDUCIA was aided by the abundance of formal variants (e.g., *fiuza, fiuzia, fuzia, huzia*), which led speakers to prefer *confiança*. Dworkin (1989) also addresses formal variation, presenting examples of morphological variation of deadjectival nouns and how substantial variation was favorable to their loss.

First documented in Spanish language texts in the thirteenth century, *aceite* (< Hispano-Ar. *azzáyt* < Ar. *zayt* 'oil') was written with considerable variation (*azeit, azeite, azeyt, azeyte*), even within the text of a single author, as in the Spanish translation of the *Kitab al-yawarih,*[32] the *Libro de los animales que cazan,* by Abraham de Toledo, a translator of Alfonso X.

Table 3.1. Distribution of *aceite* variants in *Libro de los animales que cazan*, 1250 (87 tokens)[33]

Variant	Number of tokens	Representation (%)
azeit	15	17.2
azeite	3	3.4
azeyt	64	73.6
azeyte	5	5.7

As Table 3.1 demonstrates, *azeyt* (73.6%) is by far the most popular variant used by Abraham de Toledo in the *Libro de los animales que cazan*, so much that it is employed more than all other variants combined. This is followed by *azeit* (17.2%), *azeyte* (5.7%), and *azeite* (3.4%). Perhaps the most striking observation is that the text shows a clear preference for the consonant-final forms. One plausible explanation for this tendency is that *azeyt* and *azeit* are closer transliterations of the Arabic found in the original text than vowel-final *azeyte* and *azeite*. The author of *Libro de los animales que cazan*, as a competent scholar in Arabic, would have naturally transliterated the Classical Arabic equivalent of *azzayt* into either *azeit* or *azeyt* rather than *azeite* or *azeyte* to better reflect the original Arabic. His education, in combination with the original text itself, would serve as a constant reminder of the expected Arabic pronunciation of the Spanish borrowing. An alternative explanation for the preference for *azeyt* and *azeit* in the

[32] The name of this work, *Kitab al-yawarih*, translates as 'The Book of Animals.'
[33] We found only singular variants in the text.

text is the prevalence of extreme apocope[34] in texts from the thirteenth century. As an educated individual, the author adopted certain prestigious, high register discoursal markers in vogue at the time, including a feature such as extreme apocope.[35] On the other hand, the occasional use of non-apocopated forms (i.e., *azeite* / *azeyte*) by the author may indicate instances of unintended popular speech.[36]

While Abraham de Toledo preferred the apocopated forms in the *Libro de los animales que cazan*, the works of Alfonso X present a different result. Although the Alfonsine works are many (18 works are contained in CORDE), the reader will note the relatively small number of *aceite* tokens.[37] Nevertheless, we have a few observations regarding variation of *aceite* in Alfonsine texts.

Table 3.2. Distribution of *aceite* variants in Alfonso X, 1250-1284 (19 tokens)[38]

Variant	Number of tokens	Representation (%)
azeit	0	0.0
azeite	3	15.8
azeyt	2	10.5
azeyte	14	73.7

[34] In the first wave of apocope in Spanish, final /e/ (< Lat. E) was eliminated after a single alveolar or dental consonant (i.e., /d l n ɾ s tˢ/ but not /t/; e.g., PANE > *pan*). In the second wave, extreme apocope, final /e/ was eliminated after nearly any consonant or consonant group (e.g., SEPTE > *siet*, PARTE > *part*) (Penny 2002: 58-59).
[35] The use of extreme apocope in an Arabism (as opposed to a patrimonial or French term), may suggest integration of the term in Castilian. After all, Tuten (2003) considers extreme apocope a norm of Castilian that occurred as a result of koineization.
[36] Tuten (2003: 172-173) explains that, through the end of its documentation (i.e., late thirteenth century), extreme apocope tended to dominate the written language, whereas full forms demonstrate the influence of speech or informal language.
[37] The works of Alfonso X, who generally preferred the use of Romance terms over Arabisms (García González 1993-1994: 358-359), show a strong preference for lexical variant *olio* over *aceite*.
[38] We found only singular variants in the text.

Again, as above, there is a preferred variant used by the author, *azeyte* (73.7%), which is employed more frequently than all other variants combined. Following in frequency are *azeite* (15.8%) and *azeyt* (10.5%), with *azeit* (0.0%) failing to appear at all. By dividing the variants of *aceite* into two broad groups, forms ending in <e> and those ending in <t>, we see that the former are strongly favored in these texts. Although the number of tokens is small, this observation follows the general preference in Alfonsine works for forms ending in <e> over equivalent apocopated ones.[39] While these works obviously showed some variation, Kasten (1990) notes an effort toward maintaining consistency in features such as their script and verb morphology, which in turn reveals the intention of the monarch to establish a standard to be imitated. We observe the very same effort and intention in the preference for *azeyte*, one adaptation of *azzayt* into Old Castilian. Undoubtedly, these preferences would later become widespread, made evident by the elimination of extreme apocope in his works and in that of other writers.

While historically there existed a considerable number of *aceite* variants, a small degree of variation alone was not sufficient enough to have displaced the Arabism. With orthographic variants of *aceite* in mind, one might also consider that, despite its historical variation in orthography, the word has been rather consistent in its phonology. Even under the assumption that cases of extreme apocope such as *azeit* and *azeyt* were represented phonologically as /adᶻeit/, the earliest variants of *aceite* would have had only two principal phonological forms (i.e.,

[39] Alfonsine works obviously contain some orthographic variation, with forms ending in <e> often found in the same texts as their counterparts ending in consonants. In addition, we have found that forms ending in <e> (e.g., *noche, parte*) generally enjoy considerably greater diffusion in Alfonsine texts than their apocopated equivalents (e.g., *noch, part*). The one exception that we discovered was the case of *príncipe*, with apocopated variants (i.e., *princep, príncep, prinçep, princip*) representing 446 tokens across 11 documents and non-apocopated variants (i.e., *principe, prinçipe, príncipe, prínçipe*) representing 110 tokens across 7 documents.

/adᶻeit/, /adᶻeite/), and both apocopated forms are short-lived when compared to the remaining forms, of which all are non-apocopated.[40] Furthermore, subsequent variants could only have differed in the pronunciation of their sibilant (e.g., *açeite* /atˢeite/ vs *aseite* /azeite/), if at all, since medieval and early modern orthography is not always a reliable indicator of differences in pronunciation.[41] Finally, with its one token in the documentation of the time period, we consider *aceito* as an outlier, as Table 3.3 demonstrates.

Now, if we compare the formal variation of *aceite* to other agricultural Arabisms, *aceite* is only moderate.[42] Although *alfóncigo* 'pistachio (tree); pistachio (nut)' is still considered part of the standard, it has been displaced by Romance equivalents (i.e., *pistacho* 'pistachio (fruit),' *pistachero* 'pistachio (tree)'). In addition to other factors, the greater phonological variation of *alfóncigo* compared to *aceite* may, in small part, explain the significant success and retention of *aceite*, on one hand, and the waning usage of *alfóncigo*, on the other. Still, it is important to note that agricultural Arabisms demonstrate a range of formal variation, with some showing considerable variation (e.g., *alfóncigo*) and others (e.g., *alazor* 'safflower') not.[43]

Table 3.3. Formal variants of *aceite* in all documents from Spain, 1242-1614[44]

Variant	First documentation	Tokens
aceite(*s*)	1295	1319
aceito	1527-1561	1
aceyte	1371	91

[40] Variant *azeit* is documented between 1250 and 1275 while *azeyt* is documented between 1250 and 1293.
[41] There are also the additional challenges of modern transcription: interpreting letters that are difficult to read or translating medieval characters into modern equivalents.
[42] Historical and contemporary variants of *alfóncigo* include *alfócigo*, *alfónsigo*, *alfósigo*, *alfóstiga*, *alfóstigo*, *alfoztec*, *alhócigo*, and *alhóstigo*. As the above meanings did not have an immediate equivalent in Ibero-Romance either, writers produced various orthographic solutions for the term.
[43] We found only two formal variants: *alaçor*, *alazor*.
[44] We will examine formal variants from their introduction to 1614, the year in which the last of the Moriscos were expelled from Spain.

açeite(*s*)	1406-1411	160
açeyte(*s*)	1374	31
aseite	1344	7
aseyte(*s*)	1293	74
azeit	1250	25
azeite(*s*)	1250	1138
azeyt	1250	70
azeyte(*s*)	1242 - 1275	3624

Still, by comparison, *olio* is far less variable in form than *aceite*, despite the fact that the patrimonial term is documented much earlier than its Arabic counterpart. With a total of four orthographic variants, *olio* differs only in one grapheme (i.e., <e>, <i>, <j>, <y>), representing the semivowel [i̯] (cf. Table 3.4).[45] Based on lesser formal variation alone, it is *olio* that should have been prevailed. However, it is important to recognize the multicausality of lexical change (Álvarez de Miranda 2009: 153); that is, it is rare that any one factor works alone to eliminate or retain lexical items.[46]

Table 3.4. Formal variants of *olio* in all documents from Spain, 1052-1614

Variant	**First documentation**	**Tokens**
oleo(*s*)	1052[47]	159
olio(*s*)	1196	3555
oljo	1331	39
olyo(*s*)	1200	98

In this case, despite the greater orthographic variability of *aceite* in comparison to *olio*, the Arabism was favored over its patrimonial counterpart in other ways. In other words, although

[45] Furthermore, it is possible that the <e> of variant *oleo* represented the allophone [e] as a reflection of its orthography, despite the later preference to pronounce the term as semivowel [i̯].
[46] Clearly, considerable formal variation alone was not sufficient to displace the Arabisms *truchimán*, *trujamán*, *trujimán*, etc. all of which maintain some version of the meaning of 'interpreter.'
[47] The formal variant *oleo* is first documented in 1052 in an Aragonese notarial document composed in Classical Latin. A later document showing signs of early spoken Castilian is dated 1189.

formal variation is not particularly favorable to the retention of *aceite*, there are other linguistic factors that will favor it, as we will see in the sections below.

3.2.2 Related vocabulary

Since the olive was of immense economic importance to medieval Spanish states, *aceite* (and *aceituna* 'olive') served as the base for a number of derived words, of which many suggest the direct handling of, or contact with, olives or (olive) oil.[48] Within this lexical group, we have identified four semantic subcategories: (1) cultivation and production, (2) storage and preparation, (3) trade, and (4) description or quality. Admittedly, there is some overlap of semantic categories, and several of these terms have multiple meanings.[49] In the category of cultivation and production (i.e., the individuals and resources involved in the creation of olive-based products) belong *aceitero*$_1$ 'person who sells or produces oil' and *aceituno* 'olive (tree).' The category of storage and consumption contains *aceitar*, *enaceitar* '(to) oil, (to) put oil on,' and *aceitera*$_1$ 'oil bottle, container.' In the category of trade are *aceitera*$_2$ 'business dedicated to oil,' *aceitería*$_{1,2}$ 'store where oil is sold; trade of person who sells or produces oil,' and again *aceitero*$_1$. Finally, in the category of description or quality, we have *aceitero*$_2$ 'pertaining to oil' and *aceitoso* 'oily.' While these two final terms do not necessarily suggest direct contact with the olive in contemporary speech, two items are apparent: (1) *aceitero*$_2$ initially did refer specifically to oil production, and (2) in its earliest documentations, *aceitoso* was used to compare the oil content of foods (e.g., almonds, walnuts), presumably with olive oil as the initial point of comparison.

[48] The exceptions to this tendency are lexical items that are more abstract, such as adjectives, which are often only comparative (e.g., *aceitunado* 'olive green').
[49] For our purposes, we only include definitions related, in some capacity, to olive trees, olives, and (olive) oil.

Early forms of lexical bases *aceite* and *aceituna* first appear around the same time in documentation (1242-1275 and 1250, respectively), and in under a century there already appears a wave of words derived from them. The first to appear are *aceitero*$_1$ (1335 *azeytero*) and *aceituno* (1348 *azeytunos*), terms directly related to the cultivation and production of the olive tree and its products. In a second wave just over a century later appear words involved in the storage and consumption of oil, namely *aceitar* (1471-1476 *azeitar*), *enaceitar* (1493 *enazeytados*), and *aceitera*$_1$ (1495 [*alcuza*] *azeitera*). Finally, at different points in the sixteenth century appear lexical items denoting descriptions or qualities, as in *aceitero*$_2$ (1575-1580 *aceitero*) and *aceitoso* (1513 *azeitosas* / *azeytosas*), and one dealing with trade, as in the case of *aceitera*$_2$ (1582 *azeytera*).[50]

The growing number of terms closely related to *aceite* (and *aceituna*) easily demonstrates the acceptance that *aceite* earned at an early stage of its introduction into the Spanish language. In fact, of the selected terms examined above, only *aceitería*$_{1,2}$ is no longer in regular use today, evidence of the level of integration of *aceite* and related terminology, not only for its historical importance but for its continued importance in modern quotidian language. Furthermore, we recognize, as Dworkin (2012) suggests, that there is a noticeable distinction between the usage of Latin-based and Arabic-based terminology with regard to words associated with the olive and its products. In other words, actions, objects, places, and individuals having close involvement with olives and olive oil were typically associated with Arabic language, speakers, or goods and, accordingly, were expressed with Arabic-based borrowings. This abundance of terminology for

[50] As a lexical item that is undocumented in either CORDE or CREA, *aceitería* does not form a particularly important part of the linguistic history of *aceite*.

quotidian concepts (e.g., *aceitar, aceitera*$_1$, etc.) certainly supports a popular origin for and diffusion of *aceite* (and its lexical legacy), as well as its firm establishment within the Spanish lexicon over the centuries. Below we will explain how *aceite* and its derived vocabulary contrast with semantically familiar Latin-based terminology.

Even greater in number than vocabulary related to *aceite* is that associated with or derived from *olio* and *óleo*, which include the following: *oleáceo* 'oleaceous (plant)' *oleaginosidad* 'oiliness,' *oleaginoso* 'oily,' *olear* 'anoint,' *oleario* 'oily,' *oleastro* 'oleaster,' *oleico* 'oleic (acid),' *oleícola* 'pertaining to olive cultivation,' *oleicultor* 'olive grower,' *oleicultura* 'olive cultivation,' *oleífero* 'containing oil,' *oleína* 'olein,' *oleografía* 'olegraphy,' *oleómetro* 'oleometer,' *oleorresina* 'oleoresin,' *oleosidad* 'oiliness,' *oleoso* 'oily,' *oliera* 'vessel for holy oil.' Despite this large number of individual lexical items, words related to *olio* and *óleo* are far less frequent in the documentation than those derived from *aceite* and *aceituna*. Overall, the former set of words is learned, and all have Latin (e.g., *oléaceo*) or hybrid Latin-Greek (e.g., *oleografía*) etymology. Given the prestige of literary languages like Latin and Greek, and their presence in scientific terminology throughout western languages, it is reasonable that many of these terms belong to the high register; to a large extent, this is true. Most belong to the higher register domains of (1) botany, (2) chemistry, (3) medicine, (4) religion, or (5) art, but a few refer to (6) cultivation and production as well.

Botanical terms, which are by far the most numerous, include *oleáceo, oleaginosidad, oleaginoso, oleario, oleastro, oleífero, oleorresina, oleosidad,* and *oleoso*. Terms related to chemistry include *oleico, oleína, oleómetro, oleosidad,* and *oleoso*. Certain historical medical

texts have included *oleaginosidad*, *oleaginoso*, *oleorresina*, *oleosidad*, and *oleoso*, all of which happen to be botanical terms as well. Religious and artistic terms are fewer, including *olear* and *oliera*, on one hand, and *oleografía*, on the other. These words are also accompanied by three others associated with the cultivation and production of the olive tree and its products, *oleícola*, *oleicultor*, and *oleicultura*.

By definition, one might argue that cultivation and production are more in line with the meanings associated with *aceite* and its derived terminology, rather than *olio* / *óleo*. However, despite the fact that many trade titles were Arabisms (e.g., *albéitar* 'veterinarian,' *alarife* 'arquitect'), the institutionalization of many trades has also displaced them with learned terminology (i.e., *veterinario*, *arquitecto*).[51] This historical fact perhaps best explains the duplication of these words in the semantic subcategory of cultivation and production, which was already occupied by Arabic-based terminology.

While *olio*- and *óleo*-derived terminology is undoubtedly varied, words derived from *aceite* are more semantically contained, having closer associations with the industry upon which it is based. There may also be fewer lexical items based on *aceite*, but while they have less semantic variety, they are arguably better integrated due to the quotidian nature of their meanings. Given their utility among the typical speaker, words derived from *aceite*, then, are the words that further established *aceite* as a legitimate form. If a derived word is eliminated more easily when its base word falls into disuse (Dworkin 1989: 381), then the expansion of a

[51] See Giménez-Eguíbar (2011) for a full discussion of four Arabic-Latin lexical pairs in the semantic field of trades.

semantic field through derivation should favor its retention.[52] In other words, expansion of the semantic category of olive oil to include aspects of its cultivation, production, and others is surely favorable to the preservation of *aceite*.

Another agricultural Arabism that identifies a plant-based product of commercial value is *algodón* 'cotton,' which served as the base for a number of derivatives in its semantic field: *algodonal* 'cotton field; cotton (plant),' *algodonar* '(to) fill with cotton,' *algodoncillo* 'milkweed,' *algodonero* 'belonging or relative to cotton,' *algodonosa* 'cotton grass,' *algodonoso* 'cottony.' Like *aceite*, an established semantic field of cotton and related terminology likely favored the retention of *algodón*. Many other maintained agricultural Arabisms are also supported by a small number of derived terminology (e.g., *adelfa, arrayán, jara*), but not all are (e.g., *nenúfar*). Of course, a well-established semantic category is certainly favorable, but there are also other factors that favor the maintenance of an Arabism, as we will explain below.

3.2.3 Semantic differences

While many modern Romance varieties contain reflexes of Lat. OLEUM 'oil' (e.g., It. *olio*, Fr. *huile*, Cat. *oli*), the usual term for expressing 'oil' in Modern Spanish is of Arabic origin (Sp. *aceite*, also Port. *azeite*). Spanish also inherited a related term, OSp. *olio*, which at one time was the only term to refer to oil until the introduction of *aceite*. Once introduced, the utility of the new term eventually led to the displacement of earlier *olio* entirely. While several factors were at play in this linguistic outcome, we agree with Lodares (1992) that the most influential linguistic factor is a semantic one.

[52] Dworkin (1989) contends that the elimination of *laydo* 'ugly' and *luengo* 'long' aided in the disappearance of corresponding *laydeza* 'ugliness' and *longueza* 'length.'

As a patrimonial Spanish term, OSp. *olio* is peculiar in its apparent lack of phonological development. In most cases, Latin /l/ in contact with an unstressed front vowel produced fricative /ʒ/ (typically represented as <j> or <i>) in Old Spanish (e.g., *paja* /paʒa/ < PALEA). The other possibility is that *olio* was realized as [ˈoʒo] as opposed to [oli̯o] or that [ˈoʒo] was one of two phonetic variants. Under either assumption, one may postulate a hypothetical **ojo* 'oil,' as Castro (1922) has. Although there is no written evidence of **ojo* 'oil,' a homonymic clash between **ojo* 'oil' and *ojo* 'eye' (< Lat. OCULUS) is one possible explanation for the semi-learned form of *olio*. This somewhat artificial character of the word in combination with the introduction of a semantically equivalent term (of Arabic origin) may explain the proliferation of *aceite* at the expense of *olio* (Corominas and Pascual 1980-1991: s.v. *aceite*).

On the other hand, Lodares (1992) contends that the elimination of *olio* is not at all explained by a homonymic clash between the reflexes of Lat. OCULUS and OLEUM. Before the thirteenth century, there was little use for a reflex of OLEUM or its derivatives in Spanish because oil was incredibly rare in Castile. At this time, few places in the kingdom were involved in oil production, which was achieved through the extraction of animal fat. By the first half of the thirteenth century, however, a substance known as *azeyte* (< Hispano-Ar. *azzáyt* < Classical Ar. *azzayt*) surfaced in texts during the reconquest of Andalusia, where olive cultivation and oil production had already had a significant economic role (Dufourcq and Gautier-Dalché 1983). With this consideration of the economic history of medieval Castile in mind, it is clear that there is no need to resort to homonymic clash to explain the replacement of *olio* with *aceite*; instead of a clash of homonyms, the loss of *olio* is better viewed as a clash of meaning.

In order to identify a new product, Spanish speakers adapted an existing lexical item from Hispano-Arabic (i.e., *azzáyt*) to specifically refer to oil derived from the olive. In other words, *aceite* may be considered a necessary (or cultural) borrowing since it arose from a necessity to represent in Spanish a new material reality to which an existing Romance term did not refer (Lodares 1992: 1147-1148).[53] Lodares further argues that, over time, *olio* fell into disuse because it referred only to a rare commodity that was incompatible with the new, and much more abundant, material (i.e., olive oil). As the use of *olio* diminished, the meaning of *aceite* gradually expanded in meaning to include oil produced from other materials, including oil derived from other plants and animals.[54] As such, we contend that the semantic change (i.e., extension) of *aceite* originates in regular language use as speakers (and writers) experiment with the term in new ways.[55]

Although Lodares proposes *aceite* as the immediate (and only) lexical solution for referring to olive oil, it is also not entirely accurate. Although we agree that *aceite* arose out of a perceived need to identify a particular product, there are instances in which *olio* is not as restricted in meaning as he leads us to believe; in fact, it is clear that before the appearance of *aceite* in texts, *olio* is, in various cases, not strictly limited to oil of animal origin:

(1) *Oliuas auras en todos los terminos tuyos. and non te untaras con el* ***olio***
'You will have olives in all your territories, and you will not anoint yourself with the oil'

[53] Similarly, agricultural Arabisms such as *acelga*, *azafrán*, and *nenúfar* identified products previously unknown to Hispano-Romance.
[54] Initially referring to oil derived from the olive, the meaning of *aceite* has expanded to include oils derived from other plants, as well as animals and geological material (e.g., *aceite de coco* 'coconut oil,' *aceite de ballena* 'whale oil,' *aceite mineral* 'mineral oil').
[55] This is in line with Traugott and Dasher (2003): "We see [regular] semantic change (change in code) as arising out of the pragmatic uses to which speakers or writers and addressees or readers put language…"

Fueros de Aragón, 1247

While there are cases of *olio* (before 1250) that indicate what is presumably an animal-derived fuel, there are examples (1250 on) where it does not. However, in cases where *olio* refers to its use as a fuel, it is usually difficult to determine by context alone whether it refers to oil derived from animals, plants, or a geological source.

(2) *Et si desta piedra fazen crusuelo pora quemar* ***olio****, o candela de cera*
'And if from this stone they make an oil lamp to burn oil, or a wax candle'
Lapidario, c. 1250

Even if we assume that most tokens of *olio* referring to fuel also refer to an animal-derived oil, these are still in the minority. Much more abundant in Spanish texts overall are plant-derived oils; from 1250 on, *olio* is widely used as a term to refer to oil derived from a number of plant sources.

(3) (...) *metan en ella tres destellos de* ***olyo*** *rosado o del* ***olio*** *uiolado*
'place in it three parts rose oil or violet oil'
Libro de los animales que cazan, 1250

In the first years of the documentation of *aceite*, we see precisely how inclusive a meaning *olio* has in texts such as the *Libro de los animales que cazan*. The documentation suggests that the term not only referred to a fuel source but also to other plant-based oils used for medicinal purposes. Consider the above excerpt of the *Libro de los animales que cazan*, where the adjectives are not describing the color of the oil but rather its source (e.g., rose, violet); *olio*

appears compatible with many other oils derived from plants, including almond, balsam, flax, lily, myrrh, myrtle, sesame, and even olives.

(4) (...) *deuen primeramiente apareiar los* ***olios*** *de bálssamo e de las oliuas* (...)
'they must first prepare the basalm and olive oils'
Setenario, 1252-1270

In other instances, *olio* is used to identify oil used in religious ceremonies (i.e., holy oil), sometimes accompanied by an adjective (e.g, *bendito, consagrado, santo*), but more often it is not. In cases where such an adjective is absent, this specialized meaning is instead determined by the context and/or source in which *olio* is found.

(5) *e untóles las manos con el santo* ***olio***
'and he anointed their hands with holy oil'
General estoria, Primera parte, c. 1275

(6) *e untauan muchos enfermos con* ***olio****, e eran sanos*
'they anointed many sick with oil, and they were cured'
El Nuevo Testamento según el manuscrito escurialense I-j-6, a 1260

In contrast, *aceite* is much more uniform in its meaning than *olio* in thirteenth century texts. Consider, for example, that the data available in CORDE indicate that there are no cases of *aceite* referring to a source other than olives in the *Libro de los animales que cazan*, the same text in which *olio* refers to oils derived from many different materials. With only a handful of references to green oil (e.g., *azeyt uerde*) and oil from green olives (e.g., *azeyt d'azeitunas uerdes*), there is little indication in this text that *aceite* meant anything other than 'olive oil.' This

fact aligns with the linguistic background of the translator, Abraham de Toledo, whose familiarity with the Arabic language would have allowed him to identify the difference between the (then) more restricted *aceite* (which referred only to olive oil) and the more general *olio* and translate the respective terms accordingly.

Again, while Lodares (1992) is correct in his identification of *aceite* as a cultural (or necessary) borrowing that was the result of a perceived need to identify a new product (i.e., olive oil), one detail of his explanation appears to be incorrect. Although *aceite* is initially used to refer to olive oil exclusively, *olio*, on the other hand, is not restricted to oil of animal origin, even prior to documentation of the term *aceite*. This fact is supported by the examples above that demonstrate *olio* referring to oil of botanical origin and even specifically to olive oil.[56]

Still, *aceite* does not maintain its semantic exclusivity indefinitely. Already by the end of the fourteenth century and the beginning of the fifteenth, the written record shows *aceite* with a significant expansion in its meaning, particularly in the way of plant-based oils. In *Sevillana medicina de Juan de Aviñón*, for example, we see how the author used *aceite* to refer to oil derived from a number of plants as diverse as the poppy, water lily, mandrake, and henbane.

[56] The use of *olio* to mean 'olive oil' in certain works during the mid- and late thirteenth century merits further discussion. The use of *olio* to refer to olive oil in the *Fueros de Aragón* (1247), for example, does not necessarily discredit the argument that *aceite* arose out of a perceived need to identify the new product: *olio* may be an individual solution on the part of the author to refer to olive oil before exposure to the term *aceite*. That same use of *olio* in *Setenario* may be attributed to the fact that Alfonso X favored Romance terms over Arabisms in cases of synonymy: "Esta resistencia a adoptar prestamos árabes lleva a la utilización exclusiva de «olio» en las dos obras mencionadas [*Primera crónica general I*, *General historia I*], evitando el arabismo «aceite-azeyte», pese a que en la segunda mitad del siglo XIII ya debía ser conocido el vocablo tras la invasión de gran parte de las tierras de Al-Andalus" (García González 1993-1994: 358-361). In the same section of the text, the author compares the usage of other Romance-Arabic pairs or groups in Alfonsine texts (e.g., *alogar-alquilar*, *guía-guiador-adalid*), revealing the preference of the monarch for Romance terms over their Arabic counterparts.

Evidence of its expansion in meaning is found in other texts as well, such as *Visita y consejo de médicos*, where this author uses *aceite* to refer to almond oil.

(7) *tome* (...) ***azeyte*** *de papauer y de escudete: y de mandragula and de jusquiamo:* (...) ***azeyte*** *de vayas*
'take oil of poppy and water lily, and that of mandrake and henbane, berry oil'
Sevillana medicina de Juan de Aviñón,
c. 1381-1418

(8) *sson cozidas con* ***azeyte*** *de almendras*
'they are cooked with almond oil'
Visita y consejo de médicos, a 1400

By invading a semantic territory originally occupied only by its Romance counterpart, *aceite* provides writers from the late fourteenth century on an alternative to *olio*. With two well-established terms with shared meaning, it is during this time period that we recognize *aceite* and *olio* as two variants of lexical variable 'oil.'

As variants of the same lexical variable, we would expect that *olio* and *aceite* would be found in many of the same domains during the period of study. Indeed, the two terms are found in nearly identical categories with roughly the same distribution in said categories. There are, however, a few notable differences of representation in certain categories, which we will explore in the tables below.

Table 3.5. Distribution of orthographic variant *olio* and orthographic variant *azeyte* by category, 1242-1614 (3371 and 3408 tokens, respectively)

Category	**Representation (Tokens)**	**Representation (Tokens)**
	olio	*azeyte*
Scientific Prose	62.85 (2119)	73.06 (2490)
Religious Prose	12.10 (408)	4.13 (141)
Societal Prose	7.56 (255)	9.94 (339)
Historical Prose	6.94 (234)	1.76 (60)
Legal Prose	5.54 (187)	5.83 (199)
Narrative Prose	1.63 (55)	1.61 (55)
Lyrical Verse	1.24 (42)	1.49 (51)
Didactic Prose	1.21 (41)	1.78 (61)
Narrative Verse	0.50 (17)	—
Dramatic Verse	0.38 (13)	0.14 (5)
Others	—	0.20 (7)

During the time period studied, Spanish writers heavily favored *olio* (62.85%) and *aceite* (73.06%) in the category of scientific prose, which of course includes subjects as diverse as economics, medicine, and agriculture, where terms referring to oils are a natural fit. Elsewhere, *olio* and *aceite* are represented similarly in each of the remaining categories, with two exceptions. One of the categories where these terms diverged in their use is religious prose, which was a more significant source of tokens for *olio* (12.10%) than for *aceite* (4.13%). In this particular category, it was *olio* that writers preferred to express 'holy oil' as demonstrated by its accompaniment by religious adjectives such as *benito*, *consagrado*, and *santo*. On the other hand, cases of *aceite* accompanied by these same adjectives are rare.[57]

The second case of divergence in usage between *olio* and *aceite* occurs in the category of historical prose, in which *olio* (6.94%) is considerably more popular than *aceite* (1.76%). In this instance, we attribute the preference for *olio* to the influence of Latin source material on the

[57] Variant *azeyte* appears alongside *bendito* in three cases, *santo* in one case, but it fails to appear with *consagrado*.

lexical choice of Spanish historical texts. It is well known that many early Spanish histories are translations or compilations of various older sources, particularly Latin language materials. Such is the case in works such as the *Estoria de España* and the *General Estoria*, which are based on numerous Latin texts. These two extensive texts, especially the *General Estoria*, form the greatest sources of *olio* tokens in this period; Alfonso X and company, having heavily relied on Latin source texts, most often employed *olio* as the vernacular equivalent of OLEUM.

As Spanish writers recognized a certain degree of compatibility between the two lexical variants, and through regular language use, *aceite* assumed additional meanings beyond the confines of its original meaning of 'olive oil.' Around the time period that *aceite* first demonstrated the meaning of 'holy oil' we find both an increase in the relative usage of *aceite* and a decrease in the relative usage of *olio* to refer to a substance previously reserved for *olio*. However, while *aceite* expands into religious prose and is used in this way for several centuries, it has never become common and has never replaced *olio* in this category, a position acquired by the Latinism *óleo*.

Table 3.6. Representation of orthographic variant *olio* and orthographic variant *azeyte* in religious prose by century, 1200-1799

Category	**Representation (Tokens)**	**Representation (Tokens)**
	olio	*azeyte*
Religious Prose (1200-1299)	3.78 (18 / 475)	—
Religious Prose (1300-1399)	25.98 (99 / 381)	1.05 (1 / 95)
Religious Prose (1400-1499)	13.72 (146 / 1064)	3.87 (62 / 1599)
Religious Prose (1500-1599)	10.25 (126 / 1229)	3.51 (51 / 1449)
Religious Prose (1600-1699)	12.60 (46 / 365)	5.89 (27 / 459)
Religious Prose (1700-1799)	—	—

Another pertinent case in the differentiation of use of *aceite* and *olio* is that of terminology in painting, a subcategory of Societal Prose. While they differ little in their representation within Societal Prose itself (e.g., 7.57% and 9.96%, respectively), the subcategory of painting does provide some insight into the development of Spanish terminology relevant to this particular art form. Specifically, these data provide evidence to support the fact that *aceite* never managed to establish itself among written references to painting. Disregarding the fact that there is no written evidence of *aceite* or *olio* in the realm of painting until the fifteenth century, *olio* is still represented nearly fifteen times more than *aceite* in the sixteenth century, when *aceite* was temporarily employed in this capacity. In other words, as in the category of religious prose, *aceite* does not offer much competition in the semantic category of painting. Instead, the meanings of 'oil paint' and 'oil painting' would later be incorporated into the neologism *óleo*.

Table 3.7. Representation of orthographic variant *olio* and orthographic variant *azeyte* in painting (within societal prose) by century, 1200-1799

Category	**Representation (Tokens)**	**Representation (Tokens)**
	olio	*azeyte*
Painting (1200-1299)	—	—
Painting (1300-1399)	—	—
Painting (1400-1499)	40.00 (20 / 50)	—
Painting (1500-1599)	71.19 (126 / 177)	4.87 (11 / 226)
Painting (1600-1699)	44.44 (12 / 27)	—
Painting (1700-1799)	100.00 (11 / 11)	—

The distinction that Spanish speakers established between *olio* / *aceite* and *óleo* in artistic terminology is by no means accidental. As with many other examples of Spanish language doublets, the higher register word has its origin in a prestigious literary language (e.g., Latin, Greek) whereas the popular, common words are either patrimonial (e.g., patrimonial *enebro* vs Latinism *junípero* 'juniper') or derive from contact with another language, whether

autochthonous (e.g. Basque-based *izquierda* vs Latinism *siniestra* 'left') or introduced (e.g., Arabic-based *aceite* vs Latinism *óleo*). In other words, over time, speakers began to distinguish *óleo*, for elevated use, from the quotidian uses of *olio* and *aceite*.

Such distinctions are the result of a lexical transition beginning as early as the fourteenth century, characterized, in part, by both the renovation and elimination of vocabulary (Dworkin 2004), with the goal of establishing order in the language.[58] One of the strategies employed to create order was through the addition of lexical items from well-regarded literary languages such as Latin or Greek.[59] As in other cases, speakers developed a learned term of Latin origin (i.e., *óleo*) that shared an etymon with a preexisting term (i.e., *olio*), but that would (at least, initially) contrast with it, refining its use to more elevated contexts (e.g., religion, painting). This Latinization (or re-Latinization) of the lexicon[60] may have indirectly helped to maintain *aceite*. That is, once *óleo* established itself as the term for its particular meanings,[61] *aceite* could establish itself as the term for a more general concept (i.e., 'oil').[62]

[58] Here we are referring to the process of standardization of the Spanish language, especially to the subprocesses of codification and elaboration of function, during which variation is reduced, on the one hand, and the language is further developed for use in an increasing number of domains, on the other. Some of the prominent writers involved in the standardization of the language include Nebrija (*Gramática de la lengua castellana*, 1492), Valdés (*Diálogo de la lengua*, 1535), Guadix (*Recopilación de algunos nombres arábigos*, 1593) Covarrubias (*Tesoro de la lengua castellana o española*, 1611). For a discussion of their contributions to the standardization of the language, see Penny (2001) for an overview. Giménez-Eguíbar (2016) discusses specifically attitudes toward Arabisms in the sixteenth century.

[59] In the *Obra de agricultura* by Herrera, the author proposes the infrequent term *escardadera* 'cultivator,' as a replacement for the Arabism *almocafre* 'weeder.'

[60] Harris-Northall (1999) discusses the re-Latinization of the Spanish lexicon during the late fifteenth and early sixteenth centuries.

[61] These meanings are 'oil paint,' 'oil painting,' and 'holy oil' but also the more general 'plant-based oil.'

[62] In other cases, Latinization has been detrimental to the vitality of Arabisms (e.g., *veterinario* / *albéitar*); however, there is no overtly negative social association with regard to oil. Furthermore, in the elimination of words that refer to trades (e.g., *albéitar*), there is a social component that is absent from objects (e.g., oil). In other words, the [- human] feature of oil appears to have aided in the maintenance of *aceite*. This contrasts with terms that refer to trades or individuals, with a [+ human] feature.

As a learned term, *óleo* is first recorded much later than its counterparts, in the first quarter of the fifteenth century (1419-1426) as an isolated case, only to reappear in Spanish written documentation in the sixteenth century (c. 1501).[63]

> (9) *Primeramente,* (...) *la santa traslación de los huesos del señor Sant Nicholás e del* ***óleo*** *sancto suyo*
> 'First, the holy transfer of the bones of lord Saint Nicholas and of his holy oil'
> *Documentación medieval de la iglesia catedral de León*, 1419-1426

Its first documentation, in *Documentación medieval de la iglesia catedral de León*, is obviously a high register reference; the *óleo sancto* described in the text is one of three high register meanings that *óleo* has retained all these years later in Modern Spanish.[64] As we saw in the other lexical variants, it does take time for it to establish its set of specialized meanings, but in the time period studied it shows a clear preference for the category of religious prose, far more so than *aceite* or even earlier *olio*.

Table 3.8. Distribution of variant *óleo* by category, 1242-1614 (125 tokens / 35 documents)

Category	Representation (%)	Tokens
Religious Prose	67.20	84
Historical Prose	8.00	10
Legal Prose	7.20	9
Scientific Prose	5.60	7
Narrative Prose	4.00	5
Societal Prose	2.40	3
Narrative Verse	2.40	3
Didactic Prose	1.60	2
Lyrical Verse	0.80	1
Dramatic Verse	0.80	1

[63] MSp. *óleo*, with its more restricted set of meanings, is not to be confused with *oleo*, an orthographic variant of OSp. *olio*.
[64] The high register meanings of *óleo* are 'oil paint,' 'oil painting,' and 'holy oil.'

As indicated above, the vast majority of *óleo* tokens in the selected time period occur in religious prose (67.20%), with the majority of these tokens, in turn, referring to holy oil. Naturally, this is the result of the establishment of a Latin term for oil with specialized purposes, in this case for use in the Christian sacraments.

While *óleo* is found in nearly all the same categories as *olio* and *aceite*, its distribution differs noticeably from its counterparts, especially with regard to the categories of religious prose and scientific prose. The use of *óleo* in religious texts is straightforward, given its specialized meaning, as explained above. As for its rareness in scientific prose, *óleo* was unnecessary in a category that was already saturated with two highly popular, generic terms for oil, not to mention that *óleo* does not signal a large departure in meaning from either *olio* or *aceite*. That is, the only true difference in meaning between *óleo* and the other two terms is in purpose, not in the material itself. In reality, the scope of both *olio* and *aceite* had incorporated the specialized religious uses of *óleo*; however, given the elevated status of the Christian religion in the history of the Iberian Peninsula, the idea that such a notion was worthy of its own specialized term in this regard is understandable. In other words, with the creation of the Latinism, speakers established in the semantic category of oil a distinction between the elevated (*óleo*) and the mundane (*aceite*).

3.2.3.1 Overview of semantic differences

In the time period where documentation of *olio* and *azeite* overlaps (1250 onward), we have seen how OSp. *olio* ‘oil’ was initially semantically broader than *aceite*, with an initial usage that does not present evidence of having additional meanings beyond ‘olive oil.’ During this

early period, writers were applying *olio* to oils of diverse origin, whereas loanword *aceite* was reserved for one type, olive oil, which was introduced as a product that was closely associated with the presence of Arabic speakers in the Iberian Peninsula.

Within a relatively short period of time (just over a century), when *aceite* and *olio* may be properly viewed as variants of the same lexical variable, *aceite* was first documented (1381-1418) undergoing a series of semantic changes, through regular language use, in which it was applied to oils derived from plants other than the olive. In contemporary Spanish, *aceite* is the broader of the two terms, in that it is used to refer to any oil of plant, animal, geological, or synthetic origin. The modern successor to *olio*, *óleo*, is used, generally, to refer to oils of botanical origin, but it has specialized meanings as well (i.e., 'holy oil; oil paint; oil painting').

Overall, the preservation of *aceite* alongside *olio* is greatly propelled by semantic differences that result from two opposing linguistic processes, namely extension and restriction. One explanation for the semantic restriction of *olio* is that the introduction of a cultural borrowing (e.g., *aceite*) caused an already existing term (e.g., *olio*) in the same semantic category to recoil (Bréal 1964).[65] On the other hand, social causes may explain the semantic extension of *aceite* (Meillet 1912).[66] In this example, *aceite* acquired new meanings through its use by particular social groups (e.g., physicians; medical texts are the first documented using the term to refer to plant-based oils more broadly). Alternatively, a contemporary perspective, such as that of

[65] Other examples of recoil are OEng. *fugol* 'bird' > MEng. *fowl* with the introduction of *bird* and *steorfan* 'die' > *starve* with the introduction of *die*.
[66] See Meillet (1912) for a discussion of linguistic, historical, and social causes of semantic change. Ullman (1962) proposes the additional category of psychological causes. For a contemporary perspective, see Traugott and Dasher (2003), which explains semantic change through the perspectives of historical pragmatics and discourse analysis.

Traugott and Dasher (2003), accounts for both processes: the divergent meanings of *aceite* and *olio* are the result of experimentation of speakers and writers using the terms in new ways.

As the most significant linguistic factors in the retention of *aceite,* restriction and extension worked opposite one another to maintain semantic distinctions between *olio* and *aceite*. While *olio* and *aceite* could both be said to mean 'oil' in a general sense during the majority of their individual histories, historical documentation and dictionaries reveal both subtle and obvious differences in meaning, both synchronically and diachronically. Most importantly, historical documentation demonstrates that, following the introduction of *aceite*, speakers would use the patrimonial term (i.e., *olio*) to express increasingly specialized meanings and, conversely, would use its Arabism counterpart (i.e., *aceite*) to convey increasingly broader ones.

Another agricultural Arabism, *bellota*, also competed with a patrimonial term, OSp. *lande* (< Lat. GLANS) and likewise ousted it; in this case, the patrimonial term managed to endure as *landre* following a semantic change (> 'tumor'), an example that is analogical to that of *aceite*, *olio* and *óleo*. Elsewhere, agricultural Arabisms present different outcomes, as many agricultural Arabisms, such as *adelfa*, *azafrán*, *nenúfar*, and *jara*, filled lexical voids in Hispano-Romance; in other words, there were no equivalent patrimonial terms with which to compete, and no semantic changes were necessary. In a broad sense, agricultural Arabisms have been successful due to their capacity, in many cases, to identify new concepts (e.g., *adelfa*, *azafrán*, *nenúfar*, *jara*) or, if in competition with other Romance terms, are able to distinguish themselves from their counterparts (e.g., *aceite*, *bellota*).

3.3 Sociolinguistic factors

Now that we have determined the degree to which different linguistic factors have been favorable to the retention of *aceite* in section 3.2 (formal variation, not particularly favorable; related vocabulary, favorable; semantic differences, very favorable), we may address the relevance of sociolinguistic factors, including dialect (3.3.1), register (3.3.2), gender (3.3.3), age (3.3.4), and language contact (3.3.5). In general, regional terms, differences of usage across register, gender, and age, and favorable attitudes toward the semantic category in question are favorable to the maintenance of a given Arabism. In the case of *aceite*, as we will explain below, some of these factors are favorable, while the others are unfavorable, neutral, or inconclusive.

3.3.1 Dialect

As we explained in the previous chapter, dialectal differences often play a significant role in lexical variation, which is why it is usually the first to be explored in sociolinguistic analyses. This is not only true of lexical studies as a whole but also of those concerning Arabisms in Hispano-Romance, of which many endured despite competition with synonymous Romance terminology. This influence of dialect on the preservation of Arabic loanwords is noticeable in the vocabulary of southern varieties of Castilian Spanish given the enduring presence of Arabic language in the South, on the one hand, and its earlier elimination from the North, on the other. The result is greater retention of Arabisms in the South, especially in rural varieties, compared to the North. The general tendency is that when two lexical variants of different origins exist, the Arabism is a regional term, whereas the Romance or Latin term is the standard. This is true of many competing Arabic-Romance lexical groups, particularly with regard to botanical terminology but not of *aceite*.

What is unusual about *aceite* as an agricultural or botanical term that coexists with a Romance term is that not only is this Arabism not exclusively regional, but it is not even restricted to use in Spain; instead, it is part of the standard throughout the Spanish-speaking world. This contrasts with the many botanical Arabisms that have become rare in use even in Spain. Some of the botanical Arabisms that coexist in (mainly Peninsular) Spanish alongside Romance counterparts are *almoraduj* 'marjoram; sandalwood' (vs *mejorana* 'marjoram'), *alhucema* 'lavender (plant)' (vs *espliego* 'lavender (plant)' or *lavanda* 'lavender (plant); lavender perfume; lavender (color)'), and *arrayán* 'myrtle (plant)' (vs *mirto* 'myrtle (plant)' or *murta* 'myrtle (plant); myrtle (fruit)'). A few botanical terms of this type also include a number of Arabisms that are limited to the southern regions of Spain (especially Andalusia), whether as a whole or restricted to a particular area thereof (e.g., *aljuma* 'pine needle' vs *pinocha* 'pine needle').

Given the universality of a term such as *aceite* and the association of olive oil with Arabic speakers, it is not surprising that a great deal of terminology related to olive oil production also endures in modern-day Andalusia. Examples of documented vocabulary concerning oleiculture include, but are not limited to: *almazara* 'oil mill,' *almacén* 'storage compartments of an oil mill,' *alhóndiga* 'storage compartments of an oil mill,' *alfarje* 'bed stone of a mill; canal leading from the bed stone of a mill,' *atarfe* 'bed stone of a mill; canal leading from the bed stone of a mill,' *jaraíz* 'bed stone of a mill,' *cofín* 'basket for carrying fruits,' *jamila* 'amurca,' *alpechín* 'amurca; olive oil waste,' *zulaque* 'olive oil waste; cloudy wine,' *aceitones* 'olive oil waste,' *mamorra* 'olive oil waste,' *azarcón* 'vessel for transferring or

decanting oil,' *alberca* 'well for depositing oil,' *aljibe* 'well in which to collect oil,' *albejín* 'vessel for depositing oil,' *zafra* 'vessel for storing oil.' Eight of these terms (*almazara, atarfe, jaraíz, cofín, jamila, zulaque, aceitones, aljibe*) are documented in the *ALEA* as part of the speech of the eastern portion of Andalusia; four (*almacén, alhóndiga, azarcón, albejín*) appear in the western portion. Two items (*mamorra, zafra*) were documented in central Andalusia, and three (*alfarje, alpechín, alberca*) appeared throughout the entire Andalusian region. That the greatest production of olive oil in the country takes place in eastern Andalusia, where there is also the greatest number of Arabisms referring to it, offers further support of the contribution of Arabic to the lexicon of olive oil and its production (Garulo 1983: 43-47). In other words, although *aceite* was originally a regional (i.e., Andalusian) term, it spread and established itself in other territories, upon which it could no longer be considered regional. This claim is in line with Lodares, who contends that the production of olive oil was limited to southern Iberia (i.e., Andalusia) until its introduction into Castile and Aragon in the thirteenth century, after which time olive oil and the term referring to it spread to the rest of the Peninsula (1992: 1147-1148).[67]

In contrast, terminology in the area of vinification demonstrates how that of oleiculture might have been, had the economic value of the olive been delayed until centuries later. In the area of winemaking, Romance terminology dominates (e.g., *vino* 'wine' < Lat. VINUM, *vid* 'grapevine' < Lat. VITIS), and Arabisms (e.g., *alarife* 'tendril') are scarce. The few Arabisms that exist are relegated to unimportant concepts (e.g., 'cloudy wine,' 'tendril') and are only found in restricted geographic areas. Fernández-Sevilla Jiménez explains that the modern (i.e.,

[67] Refer to section 3.1.4 for an overview of our discussion of the social and economic history of the olive in medieval Iberia.

Romance) vocabulary found in the semantic category of vinification is justified in the relatively recent economic importance of the grape and wine in Andalusia (1975: 249).

3.3.2 Register

The role of register in the retention of *aceite* is perhaps best understood within the context of semantic change as it applies to *olio* and *aceite*. As we revealed in section 3.2.3 above, the principal change in *olio* in its transformation into *óleo* is one of semantic restriction, in which a word referring to any type of oil from any source became one with a number of specialized meanings. On the other hand, the development of *aceite* is one of semantic extension, in which a term referring to olive oil only became one that may refer to virtually any oil or oily substance. In this way, the individual histories of these lexical items have been effectively opposite of one another, each having moved into semantic areas that the other had abandoned. As a result, the registers in which these terms are used have also changed and are responsible, in part, for the resilience of *aceite* since its incorporation into the Spanish language.

In sections 2.4.2.1 and 2.4.2.2 above, we learned how the distribution of *olio* and *aceite* has changed from their first documentation until the expulsion of the Moriscos in the seventeenth century. What these terms have in common is clear: both are generally found in the many of the very same text categories and are favored for use in high register texts in general. Of the high register text types, it is scientific prose that predominates for each term and in each century in which they are documented, though the degree of dominance in this category varies considerably from one century to another, especially in the case of *aceite*. For the most part, identifying a distinction between high and popular registers based on text type is more complicated than it

may appear; for example, virtually all scientific prose texts may be considered high register, many (but not all) legal prose texts contain examples of popular speech, and historical prose contains plenty of examples from both the high register (e.g., historiography) and popular speech (e.g., personal letters). It is for this reason that text type, while informative in its own right, must be evaluated with an examination of individual examples of *olio* and *aceite*.

3.3.2.1 Coexistence of *olio* and *azeyte*

In the thirteenth century, all tokens of *olio* and *aceite* that appear in the category of scientific prose may be considered high register examples, documented in Alfonsine works or translations of important, earlier scientific works from other languages. The representation of scientific prose is greater for *aceite* (65.28%) than for *olio* (41.59%) in the early stages of their documentation, but this preference for the former in such texts may be explained by the semantic precision of *aceite* compared to the more generic *olio* at this time. All tokens in historical prose, with greater representation for *olio* (28.85%) than *aceite* (4.17%), are also high register examples, all deriving from various historical works of Alfonso X. However, the category of legal prose is rather more informative: *olio* appears far more often in high register sources than *aceite*, whose diffusion takes place entirely through personal communication and notarial documents in the thirteenth century. In other words, although *aceite* is found in both elevated and popular registers, the Arabism is more representative of quotidian language than its Romance counterpart, placing *aceite* on the more informal side of *olio* in the continuum of formality.[68]

[68] Refer to section 2.3.4.2 for a brief explanation of the continuum of formality.

In the fourteenth century, scientific prose is still represented to a greater degree in *aceite* (53.51%) than *olio* (46.46%), but the difference in representation decreases considerably from the previous century. Although *aceite* was once a term with only one apparent meaning (i.e., 'olive oil'), the presence of a diverse group of modifiers of *aceite* (e.g., *de alacranes*, *de vayas*) in the fourteenth century indicates the beginning of the generalization of the term. With its increasing lack of precision, it is reasonable that the use of *aceite* within the domain of science gradually declines. During this time, *olio* is often accompanied by modifiers (e.g., *de laurel*, *de murta*), as in the previous century, although somewhat less frequently. Legal prose is in decline for both terms, with *aceite* (19.30%) still showing a greater representation in this category than *olio* (8.33%). In this century, legal prose is the only source of popular speech tokens for *olio* and *aceite*. The third major category that *aceite* (10.53%) and *olio* (9.09%) share during this time period is that of historical prose, of which all tokens belong to the high register. Other frequent categories for *olio* are religious prose and societal prose, with texts that all may be classified as high register.

The fifteenth century shows the representation of scientific prose in *aceite* (74.88%) and *olio* (73.10%) at its highest. Not only does scientific prose account for nearly three quarters of all tokens of either lexical item, but it is during this time period that the difference in the representation of scientific prose is at its lowest for these terms. The fifteenth century, with its abundance of translations of medical texts (originally composed in Latin), along with its Latin-Romance dictionaries, is a natural source of *olio* tokens. Surprisingly, however, *aceite* is a much larger contributor of tokens overall. The use of *aceite* as a translation to the original Latin (i.e.,

OLEUM) is evidence of the growing generalization of the term as it coexists with *olio*.[69] In the category of legal prose, *olio* tokens are largely found in high register texts, with many tokens appearing in a copy of the *Siete Partidas*.[70] In the case of *aceite*, the tokens consist of a combination of legal texts, notarial documents, and personal communication (i.e., letters), offering the most convincing evidence of how speakers actually employed these words; in the fifteenth century, documentation demonstrates that *aceite* continues to be used as the quotidian (and general) term for 'oil.'

In the sixteenth century, and for the first time ever, scientific prose has a greater representation among *olio* tokens (67.58%) than among *aceite* tokens (50.84%). Both *aceite* and *olio* enjoy greater diffusion into different domains within this category than in the previous century. Beyond medicine, tokens of both terms are found in texts on astrology, botany, and general biology. In addition, *aceite* is found in texts on agriculture, geography, industry, pharmacology, and zoology. The difference in representation of legal prose among *aceite* (9.26%) and *olio* (0.16%) is significant and at the greatest it has ever been. The majority of sixteenth-century texts that refer to 'oil' in the category of legal prose are notarial documents, and all such documents employ only *aceite* or one of its variants. In other words, *olio*, once a term of everyday language, is increasingly restricted to an elevated register, as demonstrated by its many tokens contained in the societal prose subcategory of painting, where related *óleo* takes

[69] We realize that individual author preferences are also at play here. Certain authors, such as the anonymous author of the *Arte complida de cirugía*, heavily favor *olio* and its variants (n = 270) to *aceite* and its variants (n = 67). Others, such as the anonymous author of *Gordonio*, strongly prefer *aceite* (n = 420) to *olio* (n = 94).
[70] Since several of the *partidas* are specifically devoted to religion and religious practices, we acknowledge that the representation *olio* tokens in legal documents may be exaggerated.

hold in the same century. On the other hand, *aceite* appears to remain in neutral territory since it is still found in both popular language and scientific texts.

In the seventeenth century, scientific prose has considerably greater representation for *olio* (54.52%) than for *aceite* (25.13%). While scientific prose is still the most frequent category for *aceite* in this century, it no longer comprises the majority of its tokens. However, the same category remains in the majority among *olio* tokens. The number of tokens in the category of legal prose, on the whole, has declined significantly such that representation is low for both *aceite* (3.29%) and *olio* (0.54%) in this regard. In fact, *olio* is so infrequently used in legal prose that it is no longer used in the category beyond the seventeenth century, and the last two such *olio* tokens appear in a notarial document from 1618. Representation of *aceite* in legal prose is also relatively infrequent, but it is higher than its counterpart; like *olio*, *aceite* is found primarily in notarial documents (e.g., especially inventories of businesses or individuals). The only clear difference between them is the degree of integration throughout the documentation (i.e., number of documents), with *aceite* (n = 23) at a much higher degree than *olio* (n = 1). By the seventeenth century, *aceite* surpasses *olio* in both the broadness of its meaning and its register; this versatility in both may explain the success (and retention) of *aceite* (n = 1520) to the detriment of *olio* (n = 365). In other words, as *olio* (and later *óleo*) became restricted to the formal side of the continuum of formality, *aceite* remained neutral in this regard.[71]

Another, dissimilar example of an Arabic-Romance pair distinguished by register is that of *alcuza* ‘vessel for oil’ and *vasija* [*para el aceite*] ‘vessel for oil.’ Like the case of *aceite*, *olio*,

[71] Escoriza Morera (2002) found that certain lexical items (e.g., *malo*) were neutral in their formality.

and *óleo*, the terms *alcuza* and *vasija* belonged to different registers. This latter set of terms differs from the former in that the Arabism *alcuza* (unlike *aceite*) was not widely circulated, known only to certain professional circles and used primarily by bilinguals or learned scholars, whereas *vasija* is a neutral term found along various points in the continuum of formality. However, there is a large number of agricultural Arabisms, and additional research will likely suggest that such Arabisms generally behave in a manner more similar to *aceite*, rather than *alcuza*; that is, a given agricultural Arabism is generally less formal than its Romance counterpart.[72] Consider, for example, the high register potential in *lavanda*, which is preferred over the more neutral *alhucema* (and *espliego*) in the context of perfumery.

3.3.3 Gender

As we addressed earlier, one of the challenges of a diachronic analysis of certain sociolinguistic factors, such as gender, is limited author information. One issue we encountered was that a large portion of texts from the fifteenth century and earlier are anonymous. In addition, tokens of *aceite* and especially *olio* among female writers are so rare that we are unable to identify gender as a relevant social factor for this lexical variable.

3.3.4 Age[73]

As in the case of gender, limited biographical information was an issue that we faced in our analysis of age. Once again, far too many texts dated from the fifteenth century or earlier are

[72] We admit that some agricultural Arabisms do not follow this tendency. It is also important to note that many agricultural Arabisms (e.g., *alfalfa*, *azafrán*) are not known to have had Romance competitors. As the only terms to identify their particular concepts, Arabisms without Romance rivals are found in texts representing various points along the continuum of formality and may therefore be considered neutral in terms of register.

[73] Author ages (and age ranges) were calculated using CORDE in combination with the *Diccionario Biográfico Electrónico de la Real Academia de la Historia*. Obviously, some authors have little or no biographical information available and could not be included in the data. In other cases, such information is approximated.

anonymous, offering very little useful information. That limitation, coupled with the fact that many of the age ranges of use largely overlapped for *aceite* and *olio*, has made it difficult for us to make any conclusive claim about this factor.

3.3.5 Language contact

As we explained in the previous chapter (section 2.3.4.3.3), the context of language contact (i.e., between Arabic and Hispano-Romance) provides important information for a better understanding of lexical retention. On one hand, we postulated that the way in which an Arabism was adopted into Hispano-Romance may have had impact on its retention. As the reader will recall, we classified loanwords of Arabic origin, broadly, as either direct or indirect, and believe that indirect Arabisms are more resistant to elimination from the language than direct ones. Equally as important is the consideration of the awareness in the speakers of the Arabic nature of the lexical item in question. Each of these aspects will be examined in turn below.

3.3.5.1 Transmission

Contact between Arabic and the other languages of the Iberian Peninsula was favored in cities, where the division of neighborhoods depended on the religion and ethnicities of its inhabitants. While private lives of the different groups were maintained separate, common areas where interaction would take place across demographic lines included the workplace and the market. In either case, the desire to communicate necessitated the use of terms that could be used (if only at a basic level) by all groups in contact with one another. As a semantic category that readily accepts additions, commerce was particularly receptive to change, giving rise to individual loanwords (Winet 2006: 194-195).

In medieval Christian Iberia, oil derived from various animal and plant sources (i.e., not olive oil) was a substance of little economic importance that was infrequently used. However, by the thirteenth century, Christian forces were introduced to olive oil after coming into contact with Arabs during and after the reconquest of Andalusia. Neuvonen (1941) agrees, highlighting the absence of *aceite* in twelfth-century texts and its sudden appearance in thirteenth-century ones, especially in and near Seville. This relatively late entry is also supported by the form of the term itself, which contains diphthong /ei/ as opposed to /ai/, the latter a phonological variant eliminated from certain texts by the turn of the eleventh century.[74] Within a relatively short period of time, from 1224 to 1248, the Crown of Castile experienced an economic shift during which olive oil became a product of major economic importance (Lodares 1992: 1147-1148). Given the importance of olive oil to the Castilian economy, *aceite* soon after appears in Spanish language documentation. In its earliest written appearance, in the mid-thirteenth century (1242-1275),[75] *aceite* is used in a *fuero* to enumerate one of many ways in which an excise on wine could be paid:

(10) *Et del* ***azeyte*** *de una morabera una panniella*
'And from one full measure of oil, he must pay the equivalent of one quarter pound'
Fuero de Usagre, 1242-1275[76]

[74] Between 900 and 950, diplomas from the Sahagún monastery showed cases of *-airo*, *-eiro*, and *-ero*; in the eleventh century, there were no longer any cases of *-airo* (Lapesa 1981: 164).
[75] According to the CORDE, the earliest documentation is from the *Fuero de Cáceres* (c. 1234-1275). Further investigation and conflicting information, however, suggest a later date for this text (1267 or later).
[76] Note that the first documentations of *aceite* are found in texts that are geographically associated with olive cultivation (e.g., Usagre, Badajoz).

This example of *aceite,* in a *fuero,* a type of legal document known for its reflection of plain speech, is evidence of the direct origin of this particular borrowing. Further evidence of this fact is demonstrated in the kind of contact that brought agricultural products and their corresponding words into Spanish, namely, person-to-person contact between Castilian speakers and Arabic-speaking merchants and producers of olive oil.

On the other hand, a large percentage of early *aceite* tokens appear in scholarly documents from the mid-thirteenth century, with the two most significant sources being the works of Alfonso X and a translation of an Arabic text by his interpreter, Abraham de Toledo. Although the documentation and lexical studies (e.g., Lodares 1992, Cano Aguilar 1993, García González 1993-1994) support the fact that *aceite* was initially transmitted and rooted as a result of direct contact of Spanish and Arabic speech communities, we must also acknowledge that the rapid propagation of *aceite* in learned texts, such as those of Alfonso X, is a reflection and further evidence of its establishment in the spoken register and its acceptance in the wider language community.

3.3.5.2 Speaker awareness and attitudes

In terms of its phonology, OSp. *azeite* easily conforms to Castilian norms (cf. *afeite*) following a few adaptations, including (but not limited to) the addition of final /e/. For this reason, this Arabism is not particularly revealing of its Arabic origin. However, some speakers of Spanish have long been aware that many Arabisms contain initial /a/, although often to the point of overgeneralization. This simplistic understanding of the facts persists, and the proposal of false etymologies based on initial /a/, even among the learned elite, are not all that uncommon,

especially in premodern eras. On the other hand, there are several thirteenth-century tokens of *azeit* and *azeyt*, which are arguably more conspicuously Arabic, but they do not survive past the thirteenth century, when extreme apocope disappears from the documentation. From the fourteenth century on, the documentation suggests that *aceite* only ended in /e/. If the only surviving form of the term were something like *azeit*, with word-final /t/, speakers might have been more suspicious of a non-Latin origin.[77]

Since orthographic variants of *aceite* represent phonological structures that are permissible in Spanish, phonological features alone would not have identified it as an Arabism. In terms of morphosyntax, *aceite* does contain Arabic article *a*(*l*)-, found in many, but not all, Arabic loanwords. As the reader will recall, some Arabisms do not contain the feature (e.g., *lima* 'lime') while those that do belong to one of two types, with either an unassimilated article (e.g., *alhucema* 'lavender (plant)') or an assimilated one (e.g., *arrayán* 'myrtle (plant)'). Originally an Arabic article, *a*(*l*)- lost all functionality once it was incorporated into Spanish, but it has nonetheless been a feature that has facilitated the identification of a number of Arabisms containing it. Naturally, for lexical items or variants containing *a*(*l*)-, association of the morpheme with Arabic is more obvious in those bearing an unassimilated article (even in cases of misidentification), but it is not necessarily clear in those containing an assimilated one. With this in mind, the assimilated article is not especially revealing of the Arabic origin of *aceite* either.

[77] Without access to premodern speech, we have to wonder if examples without <-e> represented the original, vernacular form of the word that only accepted /-e/ to adapt to Old Spanish syllable structure.

If phonological and morphosyntactic features were not adequately revealing as to the Arabic origin of *aceite*, it might have been confirmed in the realm of semantics. While the term *aceite* itself demonstrates no overt reference to Arabic culture or language, as in the cases of *algarabía* 'Arabic (language)' and *marroquí* 'Moroccan,' its close association with Arabic-speaking individuals is undeniable. With early Christian Spain producing little oil of any kind, the association of olive oil with Arabic culture and language may be explained by the fact that, until the thirteenth century, the northern regions of the Iberian Peninsula could not accommodate the culinary preferences of refugees fleeing the Arab-occupied South, where olive oil was abundant (Dufourcq and Gautier-Dalché 1983: 39-40, Lodares 1992: 1147, Dworkin 2012: 95-96). Once olive oil production methods had spread to other parts of the Peninsula, and access to olive oil had improved, the connection between Arabic culture and language was furthered established. Not only did *aceite* dominate early references to olive oil, but the entire semantic category of olive oil production (i.e., products, containers, individuals involved) was created with an almost entirely Arabic-based set of words.

Now that we have established that speakers had some degree of awareness of the Arabic nature of *aceite*, another important consideration are the attitudes toward the word in question. Of course, in a broad sense, Arabisms have long been disparaged by grammarians and the literary elite, but other causes are favorable to the introduction of the Arabism, as is demonstrated below.

The advances of Arabic culture over those of Christian Spain favored the proliferation and establishment of a diverse array of Arabisms in Spanish. That is, the incorporation of a

number of Arabisms into Spanish may be attributed to the many advances that Arabic culture had over the territories it occupied. In addition to military, political, and intellectual developments, perhaps the most conspicuous contributions are from its material culture (Winet 2006: 196-197), including agriculture (Lapesa 1981: 134-136). In fact, Arabic culture was so pervasive that even when al-Andalus faced decline, politically speaking (eleventh century onward), Hispano-Romance was still incorporating new Arabic loanwords (Colón 1998: 216).[78] However, despite the incorporation of new loanwords at this late stage, the pejoration of many Arabisms suggests that the traditional notion that all Arabisms are the result of the prestige of Arabic is inaccurate (García González 1993-1994: 353-355). Instead, another (linguistic) cause (i.e., perceived necessity) motivated the introduction of *aceite*. In other words, we argue that *aceite* is a cultural (or necessary) borrowing, which arose out of the need to identify a new material, olive oil.

Consider, for example, how Valdés explains that, contrary to expectations, there are Arabisms that were not only resistant to elimination but thrived to the detriment of a Romance counterpart:

> (11) (...) tiene oy la [lengua] castellana algunos vocablos y algunas maneras de dezir, es menester que entendáis cómo de la lengua aráviga ha tomado muchos vocablos; y avéis de saber que, aunque para muchas cosas de las que nombramos con vocablos arávigos tenemos vocablos latinos, el uso nos ha hecho tener por mejores los arávigos que los latinos; y de aquí es que dezimos antes alhombra que tapete, y tenemos por mejor vocablo alcrevite que piedra sufre, y **azeite** que olio, y, si mal no m'engaño, hallaréis que para solas aquellas cosas que avemos tomado de los moros no tenemos otros vocablos con que nombrarlas que los arávigos, que ellos mesmos, con las mesmas cosas, nos introduxeron (...)
>
> *Diálogo de la lengua*, 1535

[78] There are numerous examples of Arabisms documented in the late Middle Ages. Maíllo Salgado (1998) identifies 284 distinct Arabisms in his corpus of texts from 1300 to 1514.

The above paradoxical claim of Valdés (i.e., "no tenemos otros vocablos con que nombrarlas (...)" after having identified three Arabic-Romance pairs) has merit for our argument, at least for *aceite*. That is, the Arabism was introduced (and accepted) precisely because it was considered the only term applicable to olive oil, a new product that was entirely different from the oils used in earlier medieval Christian Iberia. Not only was *aceite* useful in having a precise meaning, it became entwined with a product of enormous economic potential. With its utility and ubiquity, *aceite* overtook a term (i.e., *olio*) used to identify a product that was economically insignificant and rare. The resulting high frequency of usage and acceptance by the language community (the latter indicated by its inclusion in formal written registers throughout its documentation, as well as the above evaluation of Valdés) indicate that attitudes toward *aceite* were at least neutral.

Another Arabism in the domain of agriculture that competed with non-Arabic counterparts is *alholí* 'granary, silo.' This Arabism, which was a term of regular usage in the sixteenth century, is found alongside a synonym of unknown origin, *troje* 'granary, silo,' in the first four editions of the *Obra de agricultura* by Herrera. In the fifth and sixth editions, *alholí* is replaced with *silo* (Giménez-Eguíbar 2016: 376-377), a Latinate term.[79] This substitution of *alholí* on the part of the author is a reflection of the perceived inferiority of Arabisms that led to its replacement. The implication, then, is that pervasive negative attitudes were sufficient enough to displace even well-established Arabisms, provided that there were lexical items with which to replace them.

[79] The *Diccionario de la lengua española* still identifies *silo* as having an unknown origin.

3.4 Summary

Throughout this chapter, we have examined the various linguistic and sociolinguistic factors involved in the history of *aceite*. Of these factors, semantic differences and language contact (i.e., neutral attitudes) are most favorable to both the introduction and retention of the lexical item. However, factors of secondary importance (related vocabulary, register) also favor retention of the Arabism. On the other hand, formal variation is not particularly favorable. A factor that is not significant in the retention of this particular lexical item is dialect, and the effects of gender and age on *aceite* are inconclusive.

As a cultural borrowing that was introduced despite the presence of a preexisting patrimonial term, speakers must not have considered *aceite* to be entirely equivalent in meaning to its earlier counterpart. In its early documentation, *aceite* was highly restrictive in its only meaning (i.e., 'olive oil'), in contrast to its counterpart, which is first documented as having a broader meaning of 'oil' without any such restriction with regard to its source. As olive oil gained importance and popularity, so did the term that referred to it, so much so that the Arabism experienced considerable semantic generalization within a few centuries. By the late Middle Ages and early modern period, *aceite* had already become generalized to the point that it could easily be applied to any oil, regardless of its origin. In this way, it assumed much of the scope that OSp. *olio* had once had in the earliest stages of its documentation. Around the same time, *olio* was undergoing a period of semantic restriction, during which it became increasingly restricted to a set of specialized meanings in the realms of religion and painting, even though its original scope included oil of any kind. Even with *aceite* and *olio* enduring opposing linguistic changes, there has always been some semantic overlap between the two terms, but any perfect

synonymy between them was either ephemeral or nonexistent, a fact that may have given *aceite*, an Arabism, its greatest chance at displacing *olio*, a preexisting patrimonial term.

The factor of language contact is certainly significant in the introduction and retention of Arabism *aceite*, but it is also a complicated one. While its earliest documentation is found in a notarial document and suggests an origin in popular speech in the early thirteenth century, *aceite* is also heavily used as early as the mid-thirteenth century in scientific texts, most notably in the works of Alfonso X. In other words, despite its origin and establishment in spoken language, this Arabism is quickly and fully accepted by the Spanish language community, as made evident by its rapid integration into formal texts. In terms of formal properties, there are no phonological, morphological, or syntactic features that undoubtedly reveal its Arabic origin. Although *aceite* does contain the Arabic article (in the form of /a/), initial /a/ is not necessarily indicative of Arabic origin (cf. patronymic *atar* < Lat. APTARE). Semantics, however, appears to be more revealing, as Arabic speakers are largely responsible for the reintroduction of olive oil in the Iberian Peninsula.

This association of olive oil and speakers of Arabic is well known, and it is for this reason that it is difficult to ignore the origin of *aceite* on the part of Spanish speakers. Instead, use of the Arabism continued well after language purists began to voice negative attitudes directed toward the Arabic language. Since many Arabisms were eliminated from the Spanish language simply due to their origin, rather than their redundancy, some of those that survived also have to be explained by neutral attitudes. We contend that one such case is *aceite*, which gained favor through its perceived need, capable of referring to a new, unfamiliar material (i.e.,

olive oil). It is only through this multicausality of linguistic and sociolinguistic factors that the fate of *aceite*, competing with a much older, Latin-based synonym, may be better understood.

Of course, *aceite* is one example of the many agricultural Arabisms introduced into Hispano-Romance. In terms of formal variation, *aceite* is moderate, although it demonstrates far less variation in the documentation than other agricultural Arabims such as *alfóncigo*, *alholí*,[80] or *arrayán*. On the other hand, *aceite* is one of the most developed in the domain of agriculture, with many other similar Arabisms having few (e.g., *arrayán*, *bellota*) or no (e.g., *acelga*, *nenúfar*) terms derived from them. We find it reasonable that the lexicon of oleiculture is more developed than other categories due to the greater economic importance of olive oil when compared to most other agricultural products during the time period studied. In general, many of the agricultural Arabisms are introduced to fill lexical voids, so *aceite* is unusual in that regard, given the potential for olive oil to have fit into the semantic meaning of the broader *olio*. However, the initial, exclusive application of *aceite* to olive oil suggests that that is precisely the function of this Arabism: filling a perceived lexical void of a previously unfamiliar type of oil.

As an agricultural Arabism that has a Romance synonym, *aceite* is also different from other examples (e.g., *alhucema*, *almoraduj*, *arrayán*) in its limited geographic distribution. Again, this may be due to the greater economic importance of olive oil compared to other agricultural products. Another consideration is that *aceite* and *olio* / *óleo* have never had the

[80] *Alholí* has a considerable number of documented formal variants. Here are those included in Giménez-Eguíbar (2016): *alholí, alhory, alhorín, alfolí, alorí, alborín, algorrín, alfocí, alforiz, alforín, algorín, ajorí.*

semantic overlap typical of regional agricultural Arabisms.[81] As far as register is concerned, *aceite* and *olio* / *óleo* behave like other Arabic-Romance lexical groups, with a distinction in use across register (e.g., *alcuza* vs *vasija*). Finally, *aceite* is different from other agricultural Arabisms that competed with Romance rivals in that attitudes are generally negative and lead to the displacement or elimination of the Arabic borrowing, which is evidently not the case with *aceite*, a well-established Arabism that continues to be the usual Spanish term to refer to oil.

[81] Evidence of this is found in their different treatment in the *Diccionario de la lengua española*: *alhucema*, for example, is defined as *espliego*, whereas *aceite* and *óleo* are defined in their own terms. *Olio* is defined as *óleo* in the *DLE*. Still, based on usage and form, we find that they are better considered separate terms.

4 Case study: *atalaya* and *centinela*[82]

In the present chapter, we consider the second of our Arabic-Romance lexical groups, that of *atalaya* 'sentinel; watchtower' and *centinela* 'sentinel,' two synonyms that we have identified as representative of the broader domain of the military. Through an examination of both linguistic and sociolinguistic factors, as well as the roles of the sentinel and watchtower in the Iberian Peninsula, we explain which are favorable, unfavorable, or irrelevant in retaining the Arabism (i.e., *atalaya*). We argue that the introduction of *atalaya* represents the broader development of a Hispano-Romance military lexicon in the medieval and early modern eras and that maintenance of this particular term is supported by the intensity of contact between Arabic and Hispano-Romance during the Reconquest, as well as neutral attitudes toward it. The introduction of its partial synonym, *centinela*, on the other hand, is one example of many that form part of the lexical transition of Castilian that occurred between the second half of the fourteenth century and the first half of the seventeeth. More generally, these findings will offer further explanation as to why military Arabisms were maintained despite the pervasiveness of generally negative attitudes toward the Arabic language and Arabic-based loanwords.

Below, we will briefly discuss Arabisms and Italianisms introduced into the domain of the military before focusing our attention on the sentinel and watchtower more specifically. In the following section (4.1), we describe the functions of sentinels and watchtowers in medieval Iberia, relying primarily on archeological studies of watchtowers to determine the responsibilities

[82] In addition to these two principal lexical variants, there are other closely related terms: *vigía* 'watchtower; sentinel' and *torre de vigilancia* 'watchtower.' Cases of these terms from the time period studied are exceedingly rare (i.e., one token and zero tokens, respectively, in Spanish documents in CORDE). As such, they will only be considered briefly in our study.

of the sentinels who were stationed therein. In section 4.2, we address the role of linguistic factors (i.e., formal variation, related vocabulary, semantic differences) in the maintenance of *atalaya*. In section 4.3, we analyze the impact of sociolinguistic factors (i.e., dialect, register, language contact). In the analysis, we will again use textual evidence available in CORDE (for formal variation, semantic differences, register), historical and contemporary dictionaries (related vocabulary), and lexical studies (dialect, language contact) to explain the impact that each factor has had on the Arabism. The following, and final, section of the chapter (4.4) will provide an overview of the results.

4.1 Military vocabulary in medieval and early modern Iberia

López Vallejo (2008) explains that Castilian Romance has borrowed and maintained military terminology of diverse origins, including Latinisms, Germanisms, Gallicisms, Occitanisms, Catalanisms, Arabisms, and Italianisms.[83] Prior to the sixteenth century, Latinisms were introduced most notably in the thirteenth century as writers (especially Alfonso X) developed the Castilian lexicon to address concepts and ideas across different domains, including the military. Some military Latinisms include *clavero* 'knight entrusted with the defense of a castle or convent' and *columna* 'column (military formation)' and also Greek-based Latinisms such as *espada* 'sword' and *sarisa* 'pike used by soldiers of the Macedonian phalanx.' Military borrowings that some scholars treat as Germanisms include *banda* 'group of soldiers that form a troop,' *espía* 'spy,' *guerra* 'war,' *yelmo* 'helmet,' and others; however, it is important to note that many terms that are treated as Germanisms are introduced into Castilian Romance via other

[83] As our case study examines competition between an Arabism and an Italianism, we address these in more detail than borrowings from other languages. See López Vallejo (2008) for further discussion of the contribution of military vocabulary from other languages. Dworkin (2012) discusses the broader lexical contributions of other languages to Castilian.

languages, such as Latin, French or Catalan. Gallicisms entered into Castilian Romance from the eleventh century onward, as Gallo-Romance speakers migrated in large numbers to the Iberian Peninsula. Some Gallicisms introduced in the Middle Ages are *botín* 'plunder,' *dardo* 'dart,' *flecha* 'arrow,' and *trotar* '(to) trot.' Colón (1967: 168, 180, 182) identifies *batalla* 'battle,' *maestre* 'second in command to captain in the navy,' and *palenque* 'stockade' as Occitanisms, although the similarities shared between French, Occitan, and Catalan make it difficult to determine if the borrowing is a Gallicism, Occitanism, or Catalanism (Dworkin 2012). Catalanisms are primarily introduced into Castilian between the fourteenth and sixteenth centuries. Some lexical items identified as military Catalanisms include *capitán* 'captain,' *coronel* 'colonel,' and *cuartel* 'barracks.'

Many of the military Arabisms introduced into Castilian are the result of contact between Castilian and Arabic during the major campaigns of the Reconquest, culminating in the capture of Toledo (1085) and Saragossa (1118). A significant source of language contact was the frontier between al-Andalus and Christian territories, home to a group of diverse inhabitants, including autochthonous peoples, Andalusi Christians, Christians from northern kingdoms, Berbers, and Muslims (Maíllo Salgado 2004), who acted as intermediaries between the two cultures and their respective languages. Another group that served as intermediaries between the two groups were soldiers, guides, and mercenaries, many of whom worked for both al-Andalus and Christian territories. This group of professionals is especially significant in their introduction of Arabisms, particularly as a source of military terminology.

During the first wave of Arabisms (eighth to eleventh century), few unique military Arabisms are introduced.[84] In this group are words related to military positions and fortifications (*alcaide* 'commander of troops,' *alcázar* 'fortress; palace,' *alférez* 'second lieutenant; ensign,' *atalaya* 'watchtower; sentinel') and warfare (*algara* 'raid; raider,' *annubda* 'duty of repairing castle walls for defense; call to arms,' *azaria* 'raid,' *azeipha* 'raid of Muslims in Christian territories, usually in summer'). Some military Arabisms in this first group are regularly used at the time but later fall into disuse: *alfétena* 'sedition; civil war', *annubda, azaria, azeipha*. Although *atalaya* is scarcely documented during the first wave, the term is frequently documented (and is well integrated) in thirteenth-century texts. During the second wave, (twelfth and thirteenth centuries), military Arabisms represent the second-largest group of Arabisms and identify military positions (*adalid* 'military leader; guide,' *almocadén* 'captain of troops,' *almogávar* 'mercenary raider,' *arráez* 'commander; captain; foreman in some positions'), equipment (*adarga* 'leather shield,' *alfanje* 'Moorish saber,' *almófar* 'helmet,' *velmez* 'jacket worn beneath armor'), and organization (*alarde* 'review of soldiers, parade,' *rebato* 'alarm, call to arms; uproar; commotion,' *ronda* 'guard duty or watch,' *zaga* 'rearguard').[85] These military Arabisms, which are generally more successful (i.e., better retained) than those of the first wave, are the result of long-term warfare, during which opposing factions increasingly resemble one another in both tactics and terminology (García González 2007).

Military Italianisms were generally introduced much later into Castilian, mainly in the early modern period, the result of contact between Italians and Spaniards, not in the Iberian

[84] Of the Arabisms that García González (2007) identified as belonging to the first wave of Arabic loanwords, 12 (i.e., 8.45%) of the 142 were military Arabisms.
[85] Of the Arabisms that García González (2007) identified as belonging to the second wave of Arabic loanwords, 32 (i.e., 11.90%) of the 269 were military Arabisms.

Peninsula but the Apennine Peninsula, where the Spanish military was stationed for long periods of time. While there, individuals who served as both soldiers and writers (e.g., Gonzalo de Ayora, Diego de Salazar) familiarized themselves with new warfare tactics, equipment, and fortifications, applying new terminology to concepts to which they had been exposed (Terlingen 2016).[86] Dworkin argues that the contact between Italians and Spaniards occurred among a small number of speakers, with the transmission of Italianisms (including military terms) more likely having occurred (in most cases) through the written register (2012: 151). Of the Italianisms identified by Terlingen, military terms are second in number only to terms referring to cultural life (e.g., *novela* 'novel,' *soneto* 'sonnet'). The author divides military Italianisms into the following categories, in which many terms are well integrated into Castilian: general terms (*emboscada* 'ambush'), military service (*centinela* 'sentinel'), troops (*soldado* 'soldier'), military hierarchy (*coronel* 'colonel'), equipment (*escarpe* 'piece of armor that covered the foot'), weapons (*pistola* 'pistol'), fortifications (*plataforma* 'platform').

4.1.1 The sentinel and watchtower in medieval and early modern Iberia

Due to the paucity of written accounts concerning rural medieval fortifications, there is relatively little known about the watchtower or the responsibilities of the sentinel in medieval Spain. However, archeological evidence does provide some clues as to the function of different types of towers based on structural features. Cressier (2004) broadly defines *Andalusi towers* (i.e., *torres andalusíes*), as the rural, isolated towers built in the time of al-Andalus. The author divides these into four types, each with its own principal function: *atalayas* (vigilance), *torres monumentales* (control of territory), *torres de alquería* (protection of people and goods), and

[86] Terlingen (2016) offers a more thorough description of relations between Spain and Italy, most notably (for our purposes) in the military domain. His study is also one of few that address the inventory of Italianisms in Castilian.

torres-residencias (residence).[87] For our purposes, we are primarily interested in the *atalayas* and *torres monumentales* since both constructions may be associated with *atalaya* 'sentinel; watchtower' and *centinela* 'sentinel,' as we will see shortly.

Towers with the primary role of vigilance (and a secondary role of control of territory), or *atalayas*, are identified as such by their arrangement, namely, lines or networks of towers that are circular in shape, which resulted in fortifications that were not only easier for sentinels to defend but also afforded them greater visibility. That these towers tended to be located along borders between territories, the coast, and other vulnerable areas strongly supports that these towers were used primarily for vigilance. Constructed with a sturdy base, the top floor of these towers consisted of one or several rooms, covered with a dome or simply a wooden floor (Cressier 2004: 210). Space inside watchtowers was limited, especially given their circular shape, which afforded significantly less usable space than that of rectangular or square buildings (Cabañero Subiza 1996: 151). Diurnal and nocturnal communication between watchtowers occurred through the use of smoke and light, respectively (Lampérez y Romea 1922: 217-218).

Given that many watchtowers from the various rulers and various eras of al-Andalus are still standing, the Andalusi State appears to have been active in the construction of these towers (Agüera Pedrosa, Molina Berbel and Carrillo Miras 2017: 304). Some examples include the circular watchtowers built in the province of Madrid during the Omeya dynasty (tenth century) and those built in the Vega de Granada and elsewhere during the Nasrid dynasty (fourteenth and

[87] Agüera Pedrosa, Molina Berbel and Carrillo Miras (2017: 304) and Zahran (2006: 25-26) explain that the *torre de alquería* also had a (secondary) defensive role, whereas defense was the primary function of the (*torre-*)*atalaya*. Both studies also provide several examples of both tower types found in Almería (province) and the old Kingdom of Granada, respectively.

fifteenth centuries). However, the construction of watchtowers along the coast was not as uniform, with towers built along the coast during the Moorish occupation being inconsistent in structure, most of them square or rectangular in shape. Furthermore, some of these watchtowers present a degree of ambiguity regarding their function, based on their amenities. For example, the tower of La Rijana (Granada), which is a part of a network of watchtowers, also contains a cistern and served as a refuge for a small population of fishermen (Cressier 2004: 210-211).

The *torres monumentales* are generally rectangular in shape and may also have had a secondary role in vigilance of their surrounding areas, but their primary role was military control of territory, as suggested by the constant presence of small groups of soldiers in these towers. One example of this tower type are the towers found in the present-day province of Soria with brickwork characteristic of the Omeya dynasty. Another example are the towers of Albarracín (Teruel) (ninth and tenth centuries). Based on certain Eastern architectural features, Cressier proposes two possible (conflicting) hypotheses for their origin: either these towers were built to increase the military presence of the Omeya dynasty, or they were built by local forces that opposed that dynasty (2004: 211).[88]

Originally under Moorish control, the various Andalusi towers in the Iberian Peninsula eventually fell into the hands of the Christians throughout the Reconquest. As the Christians gained control of these towers, they made them their own. In doing so, watchtowers (i.e., *atalayas*) and similar fortifications became a constant presence in the Iberian Peninsula throughout the medieval period (Gil Crespo 2013: 103-105). The omnipresence of these

[88] Zahran (2006) offers a thorough discussion of the development of fortifications in al-Andalus.

structures in medieval Iberia, as well as the superiority of Andalusi forces, likely favored the retention of some Arabisms of the first wave (i.e., eighth through eleventh century) such as *atalaya* and other, similar loanwords related to military posts and fortifications (e.g., *alcaide* 'warden,' *alcázar* 'fortress') (García González 2007: 541). After all, longer duration of contact between languages is generally favorable to borrowing (Thomason 2001: 66), and the language contact that encouraged military borrowing from Arabic during the Reconquest occurs over the course of several hundred years.

4.2 Linguistic factors

Now that we have addressed the role of the watchtower and sentinel in the medieval period (section 4.1), we will discuss the linguistic factors relevant to the maintenance of *atalaya*, even after the introduction of its Romance counterpart. Our discussion will first examine variation in form (4.2.1), then related vocabulary (4.2.2), and semantic differences (4.2.3). The subsequent section (4.3) will focus on the sociolinguistic factors responsible for the maintenance of *atalaya*.

4.2.1 Formal variation

As the reader will recall, variation in form is one of the most conspicuous linguistic factors that has the potential to impact the retention of lexical items (Dworkin 1989, Cano Aguilar 1993). In general, greater formal variation has a negative impact on retention, so limited variation should be favorable to it. Below we will examine the formal variants of *atalaya* and *centinela*, from their first documentation through the expulsion of the Moriscos (i.e., early seventeenth century).

For an Arabism that is first documented as early as the twelfth century, *atalaya* is rather restricted in its formal variation, especially when compared to other early military Arabisms (i.e., thirteenth century or earlier), such as *alférez*, with two dozen documented orthographic variants, or *anúbada* / *anúteba*, with more than three dozen.[89] Unlike our previous example of *aceite*, individual authors generally show a preference for only one variant of *atalaya*, whatever that variant may be. Only two of the authors (i.e., Alfonso X, Mateo Alemán) are documented in CORDE as having employed two variants. Below, we will examine the distribution of variants in the works of each of these authors.

Table 4.1. Distribution of *atalaya* variants in Alfonso X, c. 1270-1284 (28 tokens)

Variant	Number of tokens	Representation (%)
atalaya(*s*)	16	57.14
athalaya(*s*)	12	42.86

As demonstrated in Table 4.1, there is no obvious preference for either *atalaya* (57.14%) or *athalaya* (42.86%) in the works of Alfonso X. The difference between these two variants is clearly minimal, but the form of *athalaya* is interesting for what it may reveal about its pronunciation in the thirteenth century or the author of the texts in which it appears. Namely, the presence of digraph *th* may indicate an awareness on the part of the author that the source of *atalaya* (< Hispano-Ar. *aṭṭaláyaʕ* < Ar. *ṭalāyiʕ*) contained a sound that approximated native Spanish /t/. The nonnative sound, a pharyngealized voiceless alveolar stop (i.e., /tˤ/), could have easily been transliterated with a digraph, in the same way that the writing systems of Spanish and

[89] For the full lists of orthographic variants of *alférez* and *anúbada* / *anúteba*, consult their respective entries in Corriente 2008 (s.v. *alferes*, *anúbada*).

other languages have been modernized to accommodate unfamiliar sounds that were previously unknown to their phonological systems.[90] In any case, whether the *th* reflected the original Arabic pronunciation or not, that orthographic variant was short-lived, with such examples eliminated by the end of the thirteenth century.[91]

Table 4.2. Distribution of *atalaya* variants in Mateo Alemán, 1599-1604 (5 tokens)

Variant	Number of tokens	Representation (%)
atalaia	2	40.00
atalaya	3	60.00

As with Alfonso X, Mateo Alemán employs two variants, but there is, again, clearly insufficient data to claim preference for either *atalaya* (60.0%) or *atalaia* (40.0%). In this case, we attribute the difference in the form of the variants to variation that is common in medieval texts since both *i* and *y* were employed to represent the voiced palatal fricative /j/) (Ridruejo Alonso 1998: 729-730, Penny 2002: 64).

In short, nearly all authors in the time period examined employed orthographic variant *atalaya*, and only two authors (i.e., Alfonso X and Mateo Alemán) are documented as having employed two variants. There is insufficient data to indicate a preference of the authors for either of the variants that they employed, and therefore it is impossible to determine if usage of one term or another was conditioned by social factors.

[90] One such example are the various solutions created by writers for the palatal nasal [ɲ] in Romance: Sp. *ñ*; Gal., Port. *nh*; Cat. *ny*; Fr., It. *gn*.
[91] The only example of *athalaya* in CORDE found after the thirteenth century appears in a document based on an earlier medieval text.

In the case of *atalaya* and its variants, then, formal variation is highly limited, much more so than our earlier example, *aceite*. Specifically, we have encountered only four distinct forms recorded during the period studied, and the only major phonological difference lies in the presence or absence of assimilated Arabic article *a-*. Furthermore (and unlike *aceite*), *atalaya* shows little or no variation in form within the writing of individual authors.

Table 4.3. Formal variants of *atalaya* in all documents from Spain, 1017-1614[92]

Variant	First documentation	Tokens
atalaia(*s*)	1495	3
atalaya(*s*)	1017	576
athalaya(*s*)	1254	14
talaya(*s*)	c. 1234-1275	63

Surprisingly, Intra-Romance borrowing *centinela* (< It. *sentinella* 'sentinel') is rather variable in form, presenting at least six orthographic variants during the time period studied. Most of these variants (i.e., *çentinela*, *cintinela*, *cintinelo*, *cintinella*, *sentinela*) are infrequent, but the formal variability demonstrated by the term is much more than is expected of the early modern period, and noticeably more than that of *atalaya*.[93] Consider, for example, how the ending (*-o*) of *cintinelo* or the *ll* of *cintinella* demonstrate a clear lack of familiarity with the term. With its multiple formal variants, *centinela* is more variable, and unstable, in form than *atalaya*.

Table 4.4. Formal variants of *centinela* in all documents from Spain, 1521-1614[94]

[92] The first documentation of the formal variant *atalaya* appears in a Latin language letter preserved in a Mozarabic church in Oña (Burgos) in the eleventh century (1017) (Gómez-Moreno Martínez 1919: 125).
[93] By the first documentation of *centinela* (1521-1543), *atalaya* only has two formal variants (i.e., *atalaia*, *atalaya*).
[94] The formal variant *cintinelos* does not appear in CORDE. The only documentation that we have encountered is that of *Diálogos de contención entre la milicia y la ciencia* (1614) by Francisco Núñez Velasco. Furthermore, we will address the instability in gender assignment in terms of its semantics below, in section 4.2.3.

Variant	First documentation	Tokens
centinela(*s*)	1521-1543	555
çentinela(*s*)	c. 1540-1579	4
cintinela(*s*)	1521-1543	10
cintinelos	1614	—
cintinella(*s*)	1537	6
sentinela(*s*)	1561	2

With these facts in mind, *atalaya* clearly had the advantage over *centinela* in terms of formal variation. Not only does *atalaya* have fewer documented formal variants than *centinela*,[95] but the two formal variants of *atalaya* (e.g., *atalaia, atalaya*) represent a single phonological form (i.e., /atalaja/). This cannot be said of *centinela*, which is documented with at least six formal variants that potentially represent six different phonological forms. Therefore, we believe that the limited formal variation of *atalaya* is favorable to its retention, especially considering the substantial formal variation of many other military Arabisms (e.g., *alférez, anúbada*). Of course, this is only one of three linguistic factors that we will consider in our analysis. Below, we will continue our discussion, addressing related vocabulary and semantic differences, in turn, in the following sections.

4.2.2 Related vocabulary

Given the importance of the sentinel and watchtower in the defense of Christian (and Moorish) territories in the medieval period, *atalaya* became the base from which several lexical items were created. Words derived from *atalaya* appear to be more closely related to its meaning of ‘sentinel’ than ‘watchtower,’ and the association of some derived words with the latter meaning is clearer in some words (e.g., *atalayar*) than others (e.g., *atalayador*), as we will see

[95] There are three examples in CORDE of another formal variant, *zentinela*(*s*), in 1625, 1627, and 1707.

shortly. While some of these terms are more semantically precise than others, they all appear to describe the man himself (i.e., the sentinel) or his behavior. Perhaps the only truly distinctive feature among the words derived from *atalaya* is that they represent different parts of speech, including (1) nouns, (2) adjectives, and (3) verbs. Two words derived from *atalaya* are (like their base) also nouns; these include the abstract *atalayamiento* 'act of being a sentinel' and *atalayero* 'man who served in the Army in outposts, in order to observe and inform of the movements of the enemy.' In the category of adjectives belong *atalayador* 'acting as sentinel' and formal variant *ataleador*, while *atalayar* '(to) search land or sea from a watchtower or height, in order to report what is discovered; (to) observe or spy on the actions of others' and formal variant *atalear* are verbs.

The earliest documentation of *atalaya* occurs in a Latin language letter preserved in a Mozarabic church in Oña (Burgos) in the eleventh century (1017) (Gómez-Moreno Martínez 1919: 125),[96] but it is not until the beginning of the thirteenth century that its derivatives are documented. However, once its first derivative, noun *atalayero* (1218-c. 1250 *atalayeros* / *athalayero*), appears, a new *atalaya* derivative or variant appears every few decades or so throughout the remainder of the century. Following the appearance of *atalayero* is the verb *atalear* (c 1253) and its orthographic variant *atalayar* (a 1260). Documented at the end of the thirteenth century is adjective *ataleador* (c. 1280) and orthographic variant *atalayador* (c. 1300). Documented much later, in the late fifteenth century, (Antonio de Nebrija, 1495) is the rare deverbal noun *atalayamiento*.

[96] García González (2007: 539) explains that many Arabisms that were introduced into Romance prior to the twelfth century are found in documents written by Christians who left al-Andalus (i.e., Mozarabs).

The expansion of terminology in the semantic subcategory of surveillance through the derivation of forms based on *atalaya* is certainly suggestive of the importance of the sentinel and his responsibilities for the defense of Castilian territories. Still, what is of greater significance is what this expansion meant for the status of the original *atalaya* itself, namely the further establishment of this Arabism in Spanish (García González 2007: 540). Now, although the terminology based on *atalaya* is not as abundant as that based on certain Arabisms in other categories (e.g., *aceite* in the domain of agriculture), *atalaya* derivatives are by no means few as far as military Arabisms are concerned, and we must also consider that they are documented within a relatively short time and at an early stage in the development of Castilian. As a point of comparison, all but one of the terms based on *atalaya* appear in the documentation before any of the derivatives of *aceite*. It is also important to note that this later entry in the form of *atalayamiento* is the only one of the *atalaya*-based terms that has fallen into disuse.

In contrast to its earlier Arabic-based counterpart, we have found no record of *centinela* having given rise to any derived words, but instead a rather small number of expressions. With *centinela* first appearing in the documentation in the first quarter of the sixteenth century (1521-1543 *centinelas*, 1526-1536 *centinela*), *centinela perdida* 'sentinel who is sent, exposed, to run the campaign, to better observe the enemy, and is likely to be lost' appears just over a half-century later (1582 *centinelas perdidas*). At the turn of the century (1600) appears *hacer centinela* '(to) be on sentry duty,' followed by *centinela de vista* 'sentinel who is entrusted with the prisoner so as not to lose sight of him' another half-century later (1646). Based on their definitions and their usage to describe historical campaigns, these expressions are fairly

specialized and are, naturally, most often encountered in military contexts within the documentation.

Compared with the terminology derived from *atalaya*, the expressions based on *centinela* are fewer, with a total of one verbal expression and two compound nouns. In addition, none of the expressions based on *centinela* is especially common.[97] On the other hand, words derived from *atalaya* (with the exception of *atalayamiento*) maintained their vitality well beyond the time period studied. With a larger group of derived vocabulary, greater vitality, and much earlier introduction into Spanish, *atalaya* and its derived terms are favored historically in usage over *centinela* and its expressions. In other words, the development of a well-established group of related terminology has certainly favored *atalaya*.

Other examples of retained military Arabisms are, in many cases, supported by derived vocabulary, as in the case of *atalaya*: *algara* (e.g., *algarada* 'raid,' *algarear* '(to) vociferate,' *algarero* 'raider; noisy, talkative'), *rebato* (*arrebatamiento* 'fury; extasy,' *arrebatar* '(to) grab violently,' *arrebatoso* 'fast, sudden'). In contrast, obsolete military terms such as *alfétena*, *azaria*, *azeipha*, *annubda*, despite regular usage prior to the twelfth century, never gave rise to derived vocabulary in Castilian. Still, having a well-established semantic category is only one factor that favors *atalaya* (and other military terms like it). Further explanation regarding the maintenance of *atalaya*, as well as the motivation for the introduction of *centinela*, will be provided in the following sections.

[97] The *DLE* identifies each of the terms, *centinela perdida* and *centinela de vista*, as infrequent (i.e., *poco usado*), a fact confirmed by the small number of tokens of both terms in CORDE.

4.2.3 Semantic differences

4.2.3.1 Semantics of *atalaya*

Although the Ibero-Romance languages now contain reflexes of Italian *sentinella* 'sentinel,' for centuries the only available term for 'sentinel' in these languages was one derived from Hispano-Ar. *aṭṭaláyaʕ*. These Arabisms (Port., Gal. *atalaia*, Sp. *atalaya*, Cat. *talaia*) are still in use today, even if there are now fewer meanings (e.g., 'watchtower') than there were initially (e.g., 'watchtower; sentinel'). Hispano-Ar. *aṭṭaláyaʕ* itself derives from Ar. *ṭalāyiʕ*, the plural of *ṭalīʕah* 'outpost' (Corriente 2008: s.v. *atalaia*). Corominas and Pascual explain that the borrowing of the plural form, *ṭalāyiʕ*, was generalized in Ibero-Romance as a result of its almost exclusive use (over that of the singular form, *ṭalīʕah*) in Arabic texts. With regard to semantics, the authors contend that the difference in meaning between the Classical Arabic terms and Ibero-Romance may have been the conflation of meanings of similar words (e.g., *ṭāliʕ* 'sentinel,' *ṭālaʕ* 'observe,' *taṭállaʕ* 'observe from a height') sharing the same root (i.e., *ṭ* - *l* - ʕ) on the part of Romance speakers (1980-1991: s.v. *atalaya*).

The first of the Romance forms to be documented is *atalaya*, with an initial documentation of 1017 in a Latin text originating in Oña (Burgos) (Gómez-Moreno Martínez 1919: 125, Neuvonen 1941: s.v. *atalaya*).[98] In early Castilian, the earliest documentation of *atalaya* may best be described as a sort of 'sentinel':

(1) *Violo el* ***atalaya*** *e tanxo el esquila*
'The sentinel saw him and rang the bell'
Poema de Mio Cid, c. 1140

[98] Gómez-Moreno Martínez glosses *atalaya* in the context of the text as *reconocimiento militar* 'military reconnaissance.'

Elsewhere, and in the same century, the written documentation demonstrates that *atalaya* also included the related meaning of 'watchtower':[99]

(2) *Isti sunt moiones, silicet: a fonte que nascitur in sumo Valis Transunsse in directum usque ad Cabecam de Almenara* (...)*; et inde ad illam* ***atalayam*** (...)
'These are the boundary markers, namely: from the fountain born at the top of the Transon Valley and directly to the Head of Almenara; and from there to the watchtower'
Concesión de Alfonso VIII, 1181

The use of *atalaya* as 'watchtower' flourished in the thirteenth century, at times even appearing alongside cases of *atalaya* 'sentinel' in the same work. A few of these coappearances occur in important texts such as the *Libro de Alexandre* and the works of Alfonso X. In the *Libro de Alexandre*, *atalaya* 'sentinel' appears with two different forms of the definite article (i.e., *'l* < *el* and *la*, allomorphs of the feminine definite article). Although the grammatical gender of *atalaya* 'watchtower' is indeterminate in the text, usage of other military terms ending in *-a* (e.g., *espía, guardia*) during the medieval period suggests that both *atalaya* meanings were assigned feminine grammatical gender during this time period. As both meanings were treated as feminine in gender at this time, the intended meaning of *atalaya* is only made clear within context:

(3) *suvós' en su* ***atalaya*** *como solié seer*
'he went up into his watchtower as he was accustomed'
Libro de Alexandre, 1240-1250

[99] The existence of a prototypical meaning (i.e., 'watchtower') alongside a metonymic one (i.e., 'sentinel') for *atalaya* may be explained in the conflation of meanings of other, similar Arabic terms containing root *ṭ - l - ʕ* (Corominas and Pascual 1980-1991: s.v. *atalaya*).

(4) *Ya avién esto todo los griegos entendido,*
-ca avié 'l ***atalaya*** *echado apellido-*
'The Greeks had already understood all of this, because he had called the sentinel to arms'
Libro de Alexandre, 1240-1250

(5) *viólo la* ***atalaya*** *que sedié en otero*
'the sentinel who was on the knoll saw him'
Libro de Alexandre, 1240-1250

In the *Estoria de España,* attributed to Alfonso X, there is also considerable variation in the definite article used with *atalaya* (e.g., *la, el, ell*), but its treatment as a feminine noun is clear from gender agreement in the work:

(6) *fue con tres caualleros all* ***atalaya*** *de Narseo and prisola*
'he went with three knights to the watchtower of Narseo and captured it'
Estoria de España, c. 1270

(7) *Las* ***atalayas*** *and guardas de los moros quando lo uieron dieron grandes uozes*
'The sentinels and guards of the Moors when they saw him shouted loudly'
Estoria de España, II, 1270-1284

Another Alfonsine work, the *Siete Partidas,* also features the use of *atalaya* as both 'sentinel' and 'watchtower,' but this text is rather more revealing with regard to the meaning of the vocabulary in question. The author or authors charged with the text further specify the definition of *atalaya* on multiple occasions within the text:

(8) *Otrosi las escuchas de las* ***atalayas*** *que fuesen puestas para guardar la hueste*
'also, the night watchmen of the watchtowers who were stationed to guard the army'

Siete Partidas, 1491 [a 1265][100]

(9) ***Atalayas*** *son llamados aquellos onbres que son puestos para guardar las huestes de dia veyendo los enemigos de lexos si vinieren* (...) *E eso mismo de las escuchas que son guardas para de noche*
'Atalayas are what they call those men who are stationed to guard the army by day by watching from afar to see if enemies come. And the same applies to the escuchas, who are guards at night'
Siete Partidas, 1491 [a 1265]

(10) *las rondas que andan de fuera al pie del castillo and las* ***atalayas*** *que ponen de dia: and las escuchas de noche*
'the patrols that go around outside at the foot of the castle and the day watchmen that they have stationed during the day and the night watchmen at night'
Siete Partidas, 1491 [a 1265]

For Alfonso X and his associates, it is clear that there had been some degree of specialization of the meanings of *atalaya* 'diurnal sentinel' and *escucha* 'sentinel who advances at night toward enemies to observe their movements' in contrast to other, more generic terms, such as *guarda,* which were sometimes used to define the more specialized terms *atalaya* and *escucha* (García González 1996-1997: 137-138). Returning to the usage of *atalaya* as 'watchtower,' it is curious that, despite the obvious distinction between the schedules and mobility of the *atalaya* (i.e., a stationary, diurnal post) and the *escucha* (i.e., a mobile, nocturnal post), it appears that this distinction is irrelevant for the meaning of 'watchtower,' as it appears that both *atalayas* and *escuchas* were stationed in or deployed from what was known as an *atalaya* (see example 8 above). In all, thirteenth-century documentation demonstrates a preference for *atalaya* as

[100] The first edition of the *Siete Partidas* is published in 1491, but its composition is, of course, much earlier. The exact years of composition are disputed: PhiloBiblon offers dates of composition between 1256 and 1265. Craddock explains that the manuscripts containing chronological statements indicate that the *Siete Partidas* were composed between 1256 and 1263 (1990: 191).

'watchtower' over 'sentinel,' a preference that will become more pronounced in the following centuries.

Already by the fourteenth century, *atalaya* 'sentinel' becomes ever more infrequent in the documentation at the same time that *atalaya* 'watchtower' flourishes. It is also in this century that *atalaya* appears in what initially seems to be an expression (i.e., *hacer atalaya*; cf. *hacer sentinela*), which would extend the meaning from the concrete (i.e. 'sentinel; watchtower') to the more abstract (i.e., '(to) be on guard'). However, given the paucity of *atalaya* tokens in this context (i.e., following *hacer*) in CORDE, a literal meaning is more likely here:

(11) *E quando los moros ouieron a España, deste castillo fizieron* ***atalaya*** *contra los christianos*
'And when there were Moors in Spain, they converted the castle into a watchtower to use against the Christians'

Crónica del moro Rasis, c. 1300-1344

In the fifteenth century, *atalaya* 'sentinel' is far less frequent than *atalaya* 'watchtower,' but the ongoing vitality of both meanings is demonstrated in the frequent appearance of *torre* before *atalaya* to distinguish the watchtower from the sentinel. In other words, at this time there still appears to exist ambiguity with the term *atalaya* since a number of writers still modify *atalaya* with *torre* to distinguish between the tower and the man entrusted with guarding it. Each of the following examples demonstrates the various representations that authors employed to convey the meaning of 'watchtower' at this time (e.g., *torre del atalaya, torre de atalaya, torre atalaya*):

(12) *los moros que estavan çerca de la torre del* ***atalaya*** *salieron*

'the Moors who were near the watchtower left'
Crónica de Juan II de Castilla, 1406-1411

(13) (...) *por derribar vna torre de* ***atalaya*** *que fazía mucho daño a Antequera*
'for knocking down a watchtower that did much harm to Antequera'
Crónica del halconero de Juan II, a 1454

(14) *allí estaba una torre* ***atalaya****, en que avía ciertos moros*
'there was a watchtower in which there were certain Moors'
Crónica de Don Álvaro de Luna, c. 1453

Another example of *atalaya* provides a description of the Constable of Castile.[101] In it, we see how *atalaya* from the latter half of the fifteenth century onward is no longer exclusively reserved for diurnal vigilance. This fact will not only have implications for the usage of the term *atalaya* but also for the emerging *centinela*, as we will see in the following subsection. Furthermore, our last examples from the fifteenth century reveal the first indications of vacillation in the assignment of grammatical gender of *atalaya* 'sentinel':

(15) *era una casi de día e de noche velante* ***atalaya*** *sobre todos ellos*
'he was a vigilant sentinel[102] over all of them almost day and night'
Crónica de Don Álvaro de Luna, c. 1453

(16) *E eso mismo de las escuchas que son guardas para de noche. Ca lo que fazen los* ***atalayas*** *por vista eso han ellos de fazer por oyda.*
'And the same applies to the escuchas, who are guards at night. Because what atalayas do from sight, they have to do by hearing.'

[101] The Constable of Castile (i.e., *Condestable de Castilla*) was a title given to one of the most powerful individuals in the kingdom, second only to the King. His roles included commander of the military and representative for the King in his absence.
[102] We acknowledge the possibility that the intended meaning of *atalaya* here is instead 'watchtower,' in which case this would be an example of a metaphor in which Constable don Álvaro de Luna is likened to a watchtower.

Siete Partidas, 1491 [a 1265][103]

(17) *como las **atalayas** and las escuchas deuen fazer su ofiçio*
'how atalayas and escuchas should carry out their work'
Siete Partidas, 1491 [a 1265]

In the sixteenth century, the assignment of grammatical gender to military terms becomes increasingly unstable as writers waver between masculine and feminine gender for individuals while retaining feminine gender for their corresponding activities or instruments: *el ~ la atalaya* 'sentinel,' *la atalaya* 'watchtower'; *el ~ la centinela* 'sentinel,' *la centinela* 'sentry duty'; *el ~ la guardia* 'guard,' *la guardia* 'guard duty'). López Vallejo claims that these vacillations are often linked with semantic changes, specifically metonymic processes, in which an abstraction is extended, semantically, to include the individual to which that activity refers. Although *atalaya* had already had a metonymic meaning (i.e., 'sentinel') at the time, this metonymic process is characteristic of several military borrowings from the Late Middle Ages that demonstrate instability in grammatical gender: *lanza* 'lance' > 'lancer,' *pica* 'pike' > 'soldier armed with a pike', *ronda* 'guard' > 'guard duty.' Ultimately, motivation for the vacillation of these three terms, as well as *atalaya*, is the discrepancy between grammatical and biological gender of the terms when applied to men (2008: 96-98).

Well into the sixteenth century, *atalaya* is still used by some writers as 'sentinel,' even after the introduction of Romance equivalent *centinela* 'sentinel,' first documented in CORDE in

[103] We have included this text once more here since it demonstrates an incipient grammatical instability echoed in other, similar military loanwords in the sixteenth century and later.

the early part of the century (1521-1543). Also, as earlier, the vacillation or uncertainly of the grammatical gender of *atalaya* ‘sentinel’ persists:

(18) *las **atalayas**, y escuchas, y espías y guardas*
‘the atalayas and escuchas and spies and guards’
Repertorio universal de todas las leyes de estos reinos de Castilla, 1540 -1553

(19) *los dichos **atalayas**, y guardas, y escuchas y espías*
‘said atalayas and guards and escuchas and spies’
Repertorio universal de todas las leyes de estos reinos de Castilla, 1540 -1553

(20) *Cortés aquella noche fue **atalaya** de los suyos*
‘That night Cortés served as sentinel for his company’
Segunda parte de la Crónica general de las Indias, 1553

Although *atalaya* is still used as ‘sentinel’ in the writing of certain authors, Diego Hurtado de Mendoza comments on the currency of the term in this capacity, as well as former and contemporary equivalents:

(21) (...) ***atalaya**, a que los latinos llamavan espécula*
‘atalaya, which Latin speakers called espécula’
De la guerra de Granada, 1569-1573

(22) *Lo que agora llamamos **centinela*** (...) *llamavan nuestros españoles: de noche, escucha, y de dia, **atalaya*** (...)
‘What we now call centinela our Spaniards called: at night, escucha, and by day, atalaya’
De la guerra de Granada, 1569-1573

The first example from this text demonstrates that there was a word for 'watchtower' in Latin (i.e., ESPECULA), but its reflex must have been poorly established because the only evidence found in CORDE are two cases of a toponym, [*E*]*speja* (Salamanca). In the following example, we find the first instance in CORDE in which lexical variation of *atalaya* and *centinela* is addressed. Furthermore, this second example suggests that, at this time, *atalaya* is the traditional Castilian term for 'sentinel,' whereas *centinela* is the current one.

By the early seventeenth century, it is clear that the primary meaning of *atalaya* in the documentation is 'watchtower.' At the same time, use of the word as 'sentinel' is in decline. Evidence of this fact is found in the singular gloss (and definition) of the *atalaia* entry of the *Suplemento al Tesoro de la lengua española castellana*, written by Covarrubias:

(23) (...) *[en latín, specula, ae,* ***atalaya****]* (...) *inventaron que en algunas de estas torres avía çiertos espejos*[104]
'[in Latin, specula, ae, atalaya] ... they invented that in some of these towers there was some kind of mirror'

Suplemento al Tesoro de la lengua española castellana, c. 1611

Elsewhere, the uncertainty of the grammatical gender of *atalaya* 'sentinel' continues, a phenomenon that will remain unresolved as this meaning for the term falls into disuse during this century:

(24) *supo* (…) *repartir en puestos convenientes las* ***atalayas****, espías y* ***centinelas***
'he knew how to divide the atalayas, spies and centinelas for suitable posts'

[104] Although Lampérez y Romea (1922) provides a thorough description of the features and functions of medieval watchtowers, he does not confirm whether mirrors placed in watchtowers were used to reflect light to signal nearby sentinels.

Atalaya ‘sentinel’ is gradually used outside of military contexts as speakers begin to favor its metaphorical uses in a small number of texts in the second half of the century. In these few cases where *atalaya* does not mean ‘watchtower,’ *atalaya* is perhaps better translated as ‘guardian,’ rather than ‘sentinel’:

(25) *Es esta ave la **centinela** y **atalaya** de las demás, porque (...) da una voz con que avisa á las otras*
‘This bird is the protector and guardian of the rest because it cries out so as to warn the others’
Historia del Nuevo Mundo, 1653

(26) *pienso quedarme en una posada, disfrazado y encubierto, siendo amorosa **atalaya** de Beatriz*
‘I intend to stay at an inn, disguised and undercover, as a loving guardian of Beatriz’
El hidalgo de la Mancha, jornada primera, 1673

However, the use of *atalaya* in this capacity (i.e., ‘guardian’) is rather infrequent and ephemeral, with neither of these two related meanings (i.e., ‘sentinel, guardian’) maintaining their vitality into the following century. In other words, with its decline in usage and near elimination from the documentation, the seventeenth century effectively puts an end to the use of *atalaya* as ‘sentinel.’[105]

4.2.3.2 Semantics of *centinela*

[105] Cases of *atalaya* used as ‘sentinel’ beyond the seventeenth century are exceedingly rare. Even in those very few cases, it is often accompanied by a (near) synonym (e.g., *escucha*, *vigía*, *guardia*) for the sake of clarity, presumably given the anachronistic nature of the term by that time. Evidence of the eventual disuse of the ‘sentinel’ meaning of *atalaya* is also demonstrated in the *Diccionario de autoridades* (1726), which identifies this meaning of *atalaya* as anticuated (i.e., *voz antígua*).

As we mentioned above, the Italianism *centinela* is first documented in the early sixteenth century, with transmission likely occurring through the written language (Dworkin 2012: 151). In its second documentation in CORDE, Garcilaso de la Vega (1526-1536) uses *centinela* in a nocturnal context, as part of an expression (i.e., *hacer la noturna centinela*) to mean '(to) be on night watch':

(27) *¿Qué me dirás si* (...), *haciendo la noturna* ***centinela****, la grulla de nosotros fue engañada?*
'What will you tell me if, while on night watch, our crane was deceived?'
Poesías castellanas completas, 1526-1536

Elsewhere, there is evidence of the development of a metonymic meaning of the term. In Bartolomé de las Casas (c. 1527-1561), for example, *centinela* is equivalent to 'nocturnal sentinel.' As we stated above, in the sixteenth century, a number of military terms experience semantic changes that result from metonymic processes whereby an activity or weapon (e.g., *ronda* 'guard duty,' *lanza* 'lance') is semantically extended to an individual (*ronda* 'guard,' *lanza* 'lancer'). This appears to apply to *centinela* at this time, where '(night) watch,' through metonymy, semantically extends to include the meaning of '(nocturnal) sentinel':

(28) *no se hallaba * uno que velase de noche, * que llaman* ***centinela*** *los hombres de guerra*
'they could not find one who would keep watch at night, whom the men of war call centinela'
Historia de las Indias, c. 1527-1561

In some cases, there is vacillation in grammatical gender assignment, as occurs with *centinela*, where the individual is assigned both genders, while the feminine is selected for the activity.

What these first examples have in common is their overt references to *centinela* in nocturnal contexts during its first few decades of documentation. However, this restriction of usage is short-lived, as writers soon thereafter begin to use *centinela* in diurnal contexts. The first overt reference to diurnal vigilance using *centinela* belongs to Pedro Sarmiento de Gamboa (1580-1590). While *centinela* appears sporadically in diurnal contexts throughout the sixteenth century, these are in the minority, as the usage of *centinela* in nocturnal contexts predominates.

(29) *haciendo vigilante **centinela** de noche y de día por los indios*
'on vigilant watch night and day because of the Indians'
Los viajes al estrecho de Magallanes, 1580-1590

As the sixteenth century comes to an end, the documentation offers support that *centinela* is a functional synonym of *atalaya*, even though the 'sentinel' meaning of the latter term is well in decline at this time. The first coordinated examples of *atalaya* and *centinela* appear in Alonso de Cabrera (a 1598). Due to their appearance in a religious (i.e., non-military) text, each term appears to be used as a metaphorical 'sentinel,' better translated as 'guardian, protector':

(30) *El **atalaya** y **centinela**, el profeta Isaías*
'The guardian and protector, the prophet Isaiah'
De las consideraciones sobre todos los evangelios de la Cuaresma, a 1598

López Vallejo (2008) explains that the semantic changes occurring in technicisms (e.g., military terms) that result from the process of metaphor (and metonymy) often require minimal abstraction and only basic association.

In the seventeenth century, writers continue to vacillate between masculine and feminine gender assignment of military terms referring to individuals, such as *centinela*. Again, this phenomenon is often the result of semantic changes, that is, metonymic processes whereby an activity or weapon (e.g., *ronda*, *lanza*) is semantically extended to an individual. Examples of this vacillation in the seventeenth century include *el ~ la centinela* and other similar forms but also the overtly masculine *cintinelo* provided by Francisco Núñez Velasco, which is understandable if we consider the biological gender of the referent:

(31) *vno de los* ***cintinelos*** *de dentro*
'one of the sentinels inside'
Diálogos de contención entre la milicia y la ciencia, 1614

Also, at this time, the compatibility of *atalaya* and *centinela* is further confirmed (through the end of the time period of interest and beyond) in Alonso Fernández de Avellaneda (1614) and Bernabé Cobo (1653). The former author describes the act of vigilance (*centinela*) in terms of the verb *atalayar*, whereas the latter author coordinates the terms, using them as synonyms to personify, at a time when usage of *atalaya* as 'sentinel, guardian' is in decline:

(32) *cavalleros, que* (...) *atalayáys, puestos en perpetua* ***centinela****, días y noches*
'knights, who guard, on perpetual watch, days and nights'
Don Quijote de la Mancha, 1614

(33) *Es esta ave la* ***centinela*** *y* ***atalaya*** *de las demás, porque* (...) *da una voz con que avisa á las otras*
'This bird is the protector and guardian of the rest because it cries out so as to warn the others'
Historia del Nuevo Mundo, 1653

With the introduction of *centinela* in the early sixteenth century, writers are given an alternative to earlier *atalaya* 'sentinel, guardian.' As terms with shared meanings throughout much of the sixteenth and seventeenth centuries, it is clear that *atalaya* and *centinela*, for a time, are two variants of lexical variable 'sentinel, guardian.' The motivation for the incorporation of this military Italianism may be attributed to the greater clarity offered by the term *centinela* when compared to *atalaya*, which referred to both the watchtower and the sentinel stationed therein. In other words, *centinela* is less susceptible to ambiguity, which favored its adoption (and later success) to the detriment and downfall of the 'sentinel, guardian' meaning of *atalaya* (López Vallejo 2006: 358-359). The example of *atalaya* and *centinela* closely resembles another example outside of Castilian: In Swiss Italian, *corona* meant both 'crown' and 'wreath.' Contact with German, however, led speakers to further differentiate the concepts, incorporating a borrowing, *Kranz*, for the meaning of 'wreath' and retaining *corona* for 'crown.' Weinreich explains that, in contact with another language, speakers in the target language may be motivated to borrow words into semantic fields perceived to be insufficiently differentiated (1968: 59).

4.2.3.3 Overview of semantics of *atalaya* and *centinela*

By reviewing the documentation, we have seen how *atalaya* initially contained both a prototypical meaning (i.e., 'watchtower') and a metonymic one (i.e., 'diurnal sentinel'), likely the result of the conflation of meanings of Arabic words containing the same root (i.e., *ṭ - l - ʕ*). During the fifteenth century, *atalaya* underwent semantic extension with regard to its first meaning, as demonstrated by its usage in both diurnal and nocturnal contexts, becoming simply 'sentinel.' Further extension of the 'sentinel' meaning, through metaphor, led to the addition of the meaning of 'guardian, protector.' The trajectory of *centinela* is comparable to that of *atalaya*

'sentinel,' although *centinela* appears to have developed its metonymic meaning rather than inheriting it (López Vallejo 2008: 151). Just as *atalaya* came to be used in nocturnal contexts, *centinela* came to be used in diurnal contexts. Likewise, its 'sentinel' meaning extends, through metaphor, to include the meanings of 'guardian, protector.' We argue that the semantic extension that each term for 'sentinel' experiences is regular, the result of writers (and speakers) experimenting with the terms in new ways. Some regular semantic changes that arise from experimentation of usage are semantic extension, restriction, metaphor, and metonymy, of which the last two processes require only basic associations between similar (metaphor) and contiguous (metonymy) concepts. In contrast, irregular semantic changes arise from the susceptibility of words (especially nouns) to extralinguistic factors, such as changes in referent (e.g., Sp. *coche* 'horse-drawn carriage' > 'car,' Eng. *phone* 'rotary dial telephone' > 'smartphone') or in social construction (e.g., reclamation of Eng. *Yankee*, originally a pejorative term). A change of referent occurs when a material reality changes without a corresponding change in the word that identifies it. A change of social construction may occur from an institutional intervention (AmEng. *harassment* 'annoyance' > 'aggressive pressure or intimidation'); in other cases, there is reclamation of an initially negative term (e.g., Eng. *queer*) by the community to which it refers (Traugott and Dasher 2003: 3-4).

Based on our examination of *atalaya* and *centinela*, what appears to have occurred is a restructuring of the semantic subcategory of vigilance in Castilian Romance, which consists, in part of the following: the initial introduction of *atalaya* 'watchtower; sentinel' (2) the subsequent introduction of *centinela* 'sentinel', and (3) the loss of *atalaya* 'sentinel.' First, *atalaya*, meaning both 'sentinel' and 'watchtower' in its earliest documentation in Castilian, was borrowed at a

time when speakers perceived a need to further differentiate terminology in this domain.[106] This claim would coincide with the factor that Weinreich (1968) describes as insufficient differentiation of terminology, in this case, with regard to military posts and fortifications.

Later, throughout the second half of the fourteenth through the first half of seventeenth century, writers made efforts to renovate the Castilian lexicon across various domains through the addition of loanwords, especially from Latin. Since Latin loanwords were not as polysemic as terms belonging to the popular register, this fact may have motivated the introduction of Latinisms (e.g., *ejército* 'army') despite the presence of preexisting equivalent terms (e.g., *hueste* 'army') (Dworkin 2006: 61-63, 66-67). The same may be true of military terms; however, these borrowings are most often from other Romance languages (e.g., *atalaya* vs Italianism *centinela*, *peón* 'foot soldier' vs Italianism *infante* 'foot soldier'). In other words, an explanation for the introduction of *centinela* may be what Dworkin (1989) termed excessive semantic weight, in this case, of *atalaya*, which has been recorded with both figurative and literal meanings that range from men, to heights, and to positions or states. *Centinela*, upon its introduction (and to this day), is far less ambiguous (i.e., 'guard duty' or 'sentinel'). Eberenz (2004), for his part, analyzed numerous competing synonyms from the fifteenth century and explains that the lexical transition of Castilian is motivated by the diversification of the Castilian lexicon, with the following objectives: (1) to create more precise intellectual vocabulary, (2) to expand upon synonyms for esthetic purposes, and (3) to create lexical options intended for certain text genres.

[106] Note the absence of a maintained (non-toponymic) reflex for Lat. SPECULA 'watchtower' or VIGIL 'sentinel' in Spanish. None of the various solutions for SPECULA or VIGIL match their Latin predecessors in name or, at times, in meaning (López Vallejo 2006: 361-362).

Although *atalaya* was the earliest and most successful word for 'watchtower,' in competition with the more semantically precise Italianism *centinela* 'sentinel,' *atalaya* lost the meaning it shared with its Italian counterpart. This loss of meaning in *atalaya* appears to be a case of recoil, whereby a newer term (e.g., *centinela*) causes an older one (e.g., *atalaya*) to lose its range of meaning. Other, similar cases of recoil are OEng. *stōl* 'throne' > MEng. *stool* with the introduction of *throne* and OEng. *dēor* 'animal' > MEng. *deer* with the introduction of *animal* (McMahon 1995: 176-177). Other examples of semantic restriction, in Castilian, include OSp. *arienço* 'medieval coin' < Lat. ARGENTEUS 'of silver' (cf. Sp. *argénteo* 'of silver') and Sp. *rezar* '(to) pray' < OSp. *rezar* '(to) recite' < Lat. RECITARE '(to) recite' (cf. Sp. *recitar* '(to) recite').[107] A potential motivation for the loss of *atalaya* 'sentinel' (as opposed to *atalaya* 'watchtower')[108] is its [+ human] feature, since Arabisms with a [+ human] feature tend to be susceptible to loss amid Romance counterparts (e.g., *albéitar* vs *veterinario*, *alfayate* 'tailor' vs *sastre* 'tailor'; cf. Giménez-Eguíbar 2011).

Overall, the maintenance of the term *atalaya* itself (despite its partial synonymy with *centinela*) is aided by the retention of its one distinctive meaning ('watchtower'). With both terms having endured similar types of semantic extension with regard to their 'sentinel' meanings (*atalaya* 'diurnal sentinel' > 'sentinel' > 'guardian, sentinel'; *centinela* 'nocturnal sentinel' > 'sentinel' > 'guardian, sentinel'), the end result is very much the same for *centinela* and the secondary meaning of *atalaya*, that is, close synonymy. Usually, the existence of two words with similar meanings is unnecessary, as such redundancy is generally unfavorable to the

[107] Penny provides many more examples of semantic change in Castilian (2002: 302-317).
[108] The scarce representation of the neologism *torre de vigilancia* 'watchtower' and the Portuguese borrowing *vigía* 'watchtower; sentinel' in the documentation suggests a continued preference for *atalaya* as 'watchtower.'

preservation of an Arabism in competition with patrimonial or Romance terms (cf. *azogue* 'mercury' vs *mercurio* 'mercury' or *ajebe* 'alum' vs *alumbre* 'alum').[109] However, the fact that *atalaya* had already had a unique meaning ('watchtower'), which it never shared with *centinela*, is very favorable to its retention.[110]

4.3 Sociolinguistic factors[111]

Having addressed the impact of linguistic factors on the preservation of *atalaya* in section 4.2 (formal variation, favorable; related vocabulary, favorable; semantic differences, very favorable), we may turn our attention to sociolinguistic factors, specifically dialect (4.3.1), register (4.3.2), and language contact (4.3.3). Most often, regionalized usage of vocabulary, register usage differences, and positive (or neutral) attitudes toward a given semantic category favor the maintenance of Arabisms. In the case of *atalaya*, as we will discuss below, one of these factors is favorable, another is inconclusive, and one is neutral with regard to retention.

4.3.1 Dialect

Now that we have examined linguistic factors, we may turn our attention to the relevance of sociolinguistic factors in the retention of *atalaya*. As we stated in the previous chapter, differences in dialect often play a significant role in sociolinguistic analyses, a fact that is especially true when analyzing words of Arabic origin in Spanish, which often exist only in certain regions of the Iberian Peninsula. Again, numerous Arabisms are restricted to certain parts

[109] Dworkin (2004) offers various examples of Arabisms replaced with Romance synonyms.
[110] The use of *atalaya* as 'watchtower' is well established as early as the thirteenth century. Despite the emergence of *vigía* (c. 1444) and *torre de vigilancia*, these Romance terms have never been serious competitors of *atalaya* 'watchtower.'
[111] Due to the lack of information available in CORDE with regard to biographical data (especially gender and age), we will not address the factors of age or gender in this or the following chapter.

of southern Spain due to the longer duration of contact between Arabic and Romance speakers in this part of the country. Often when there are two lexical variants of distinct origins, the Romance or Latin-based term is considered standard, whereas the Arabism exists as a regionalism (Maíllo Salgado 1998: 504). However, unlike a number of Arabisms of the semantic categories of botany and agriculture, military Arabisms with Romance or Latin-based counterparts generally do not exist as regional terms but are instead used throughout the entirety of Spain.

Like *aceite*, *atalaya* is not only used throughout Spain, but it is also employed throughout the Spanish-speaking world. As such, it behaves similarly to a number of other Arabisms in the realm of military terminology, as we will see shortly. Furthermore, we place *atalaya* in the subcategory of posts and classes of soldiers (Carrasco Cantos 1992), a specialized group of words belonging to a category (i.e., the military) that already contains its share of technicisms. Many of the words in this subcategory are featured in the works of Alfonso X, especially the *Siete Partidas*, where they are defined. In addition to defining *atalaya* 'diurnal sentinel,' this text defines several other posts borrowed from Hispano-Arabic, including *adalid*, *alférez*, *almocadén*, and *ronda*, to name a few. What *atalaya* and these other terms for military posts have in common is the significant semantic change they have all experienced through their passage from Arabic to Romance: *alférez* 'second lieutenant; ensign' (< Hispano-Ar. *alfáris* 'horseman' < Ar. *fāris*), *ronda* 'guard duty or watch' (< Hispano-Ar. *arrútba* < Ar. *rutbah* 'grade, rank'), *almocadén* 'captain of troops' < Hispano-Ar. *almuqadám* < Ar. *muqaddam* 'appointed to an office,' *adalid* 'guide, leader' < Hispano-Ar. *addalíl* 'guide' < Ar. *dalīl*. Here are the definitions of *alférez*, *ronda*, *almocadén*, and *adalid* from the *Siete Partidas*, respectively:

(34) (...) *maestro de caualleria que quiere tanto dezir como que es puesto por cabdillo: o por maestro de los caualleros del enperador: o del rey a que llaman en romançe* ***alferez***
'master of cavalry, which means that he is selected as commander, or as master of the knights of the emperor, or of the king, which they call alferez in Romance'

(35) *las* ***rondas*** *que andan de fuera al pie del castillo*
'the patrols that go around outside at the foot of the castle'

(36) ***Almocadenes*** *llaman agora a los que antiguamente solian llamar cabdillos de las peonadas*
'What they now call almocadenes they used to call commanders of the troops'

(37) *Aposentar huestes es muy grand maestria and ha menester de ser muy sabidor el cabdillo que lo ha de fazer. E para esto deuen sienpre traer consigo onbres que sepan bien la tierra a que llaman agora* ***adalides*** *que solian antiguamente auer nonbre guardadores*
'Lodging armies is a great skill, and it is necessary that the commander who must do so is very knowledgeable. And for this they should always bring with them men who know the land well, and whom they now call adalides, whom they used to call guardadores'
Siete Partidas, 1491 [a 1265]

Carrasco Cantos explains that the *Siete Partidas* are an example of the desire of Alfonso X to develop Castilian for usage in a domain (in this case, legal prose) that was previously reserved only for Latin. Using Latin, Greek, and Arabic as sources for borrowings, the monarch establishes precise definitions for a large number of terms to be used in Castilian, whether the terms were borrowings from other languages or merely terms unfamiliar to its readers. The hallmark of this text are its definitions, which are not typical of other texts from the same time period. As a legal text, the *Siete Partidas* imposed definitions on words that were intentionally precise in meaning (1992: 7-8). In many (but not all) cases, as above, the definitions of military terms provided by Alfonso X align with specialized meanings that resulted from their borrowing

into Castilian Romance as Christians adopted particular aspects of the Arab armies. Other examples include *almazen* ‘ammunition and supplies for war’ (in general usage, ‘store’) and *almoneda* ‘market of goods seized in war’ (in general usage, ‘auction’) (García González 1996-1997: 138-139).

Many Arabisms referring to military posts spread extensively, not only throughout Castilian but also to other Ibero-Romance languages. Consider, for example, how each of these terms has been established in various Ibero-Romance varieties: (1) Port., Cat. *alferes*, Gal., Sp. *alferez*, (2) Port., Gal., Sp. *ronda*, (3) Port. *almocadém*, Gal., Sp., Arag. *almocadén*, Cat. *almugatèn*, (4) Port. *adaíl*, Gal. *adail*, Sp. *adalid*, Arag. *adelid*, Cat. *adalil*, (5) Port., Gal. *atalaia*, Sp. *atalaya*, Cat. *talaia*. Given the success of *atalaya* and other terms in its subcategory in the Ibero-Romance languages, dialectal differences cannot be considered a factor relevant in the retention of this particular Arabism or for other lexical items in its subclass.

Given their success in Ibero-Romance, it may also be of interest to compare the above terms with the names of the same military posts in other Romance languages: (1) Fr. *sous-lieutenant* ‘second lieutenant,’ *enseigne* ‘ensign’; It. *sottotenente* ‘second lieutenant,’ *portabandiera* ‘ensign,’ (2) Fr. *ronde* ‘guard duty or watch’; It. *ronda* ‘guard duty or watch,’ (3) Fr. *capitaine d’infanterie* ‘captain of troops’; It. *capitano di fanteria* ‘captain of troops,’ (4) Fr. *guide* / *conducteur* ‘guide, leader’; It. *guida* / *conduttore* ‘guide, leader,’ (5) Fr. *mirador* ‘watchtower,’ *sentinelle* ‘sentinel’ ; It. *torre di guardia* ‘watchtower,’ *sentinella* ‘sentinel.’ With the exceptions of Fr. *ronde* and It. *ronda*, military posts in French and Italian are primarily assigned Romance terms. Although the above selection of terminology is small, the comparison

of Ibero-Romance military posts with those of French and Italian reflects the greater intensity of contact and military conflict in Iberia than in other territories.

4.3.2 Register

As in the previous chapter, we will consider the role of register in the maintenance of *atalaya* and *centinela* as inextricably linked to the meanings of the lexical items in question. As we stated in section 4.2.3, the principal change that both *atalaya* and *centinela* undergo is semantic extension (in their 'sentinel' meaning), with each word becoming broader in meaning over the course of its usage. This extension in meaning, of course, may have implications for the text types, and register, in which these terms appear. Notable differences in register usage between these synonymous terms should be favorable, given the tendency for speakers to reserve certain words for different points along the continuum of formality (Escoriza Morera 2002). Let us examine below the text types in which *atalaya* and *centinela* appear in order to determine the extent to which register has had an effect on the survival of *atalaya* in Spanish.

4.3.2.1 Distribution of *atalaya* prior to the introduction of *centinela*

During its first centuries of documentation (i.e., through the end of the fourteenth century) *atalaya* is restricted to a small number of text categories, namely legal prose, historical prose, narrative verse, and narrative prose. It is only in the fifteenth century and later that the term is found in a variety of text types, including the aforementioned ones and the addition of religious prose, scientific prose, lyrical verse, and societal prose. Of course, we are most interested in how *atalaya* and *centinela* behave in terms of the registers in which they are found while in competition with one another, which we will consider in the following subsection.

4.3.2.2 Coexistence of *atalaya* and *centinela*

In the sixteenth century, when they are first in competition with one another, *atalaya* and *centinela* are found in nearly identical text types, although *atalaya* tokens are found in one category in which *centinela* does not appear (i.e., dramatic prose). Unlike the early documentation of *aceite* and *olio*, there is no strongly favored category for either *atalaya* or *centinela*, suggestive of the broader (non-military) meanings (e.g., 'guardian, protector') permitted by the latter terms (i.e., through semantic extension and metaphor). During this time, *atalaya* favors scientific prose (22.98%), religious prose (20.50%), narrative prose (18.63%), and historical prose (14.91%), but other categories are represented as well, although to a lesser extent. Representation by text type is somewhat different for *centinela*, which is most frequently used in societal prose (20.28%) and lyrical verse (20.28%) but also in narrative verse (13.99%), scientific prose (13.29%), and historical prose (11.89%). In any case, representation of tokens by text type is notably more diffuse for *atalaya* and *centinela* than it is for *olio* and *aceite*. Given their representation in these text categories, both *atalaya* and *centinela* are favored in higher register texts, a fact confirmed by our own examination of text subcategories and token examples. However, a small percentage of tokens of both *atalaya* (13.04%) and *centinela* (4.90%) reflect speech or popular language. If these percentages are any indication of actual register preferences, *centinela* was preferred for higher register language at this time, which is reasonable if *centinela* was indeed transmitted via the written register, as Dworkin (2012: 151) and Terlingen (2016) suggest. Still, the usage of both terms in nearly identical text types (and sometimes the same texts) makes it difficult to make definitive claims regarding differences in usage of the terms based on register.

In the seventeenth century, *atalaya* and *centinela* are again found in nearly identical text types, with *centinela* used in one category in which *atalaya* is not (i.e., societal prose). Usage of *atalaya* is, overall, hardly different from that of the previous century, with scientific prose (24.72%) still dominant, followed by historical prose (16.29%), narrative prose (12.92%), dramatic verse (11.80%), narrative verse (10.67%), and many of the same categories from the sixteenth century.[112] On the other hand, *centinela* is best represented in narrative prose (30.04%) and historical prose (20.95%), followed by lyrical verse (13.44%), narrative verse (10.28%), and other, less frequent text types. At this time, both terms are frequent in the same categories (i.e., historical prose, narrative prose, and narrative verse) and often in the same texts. The only noteworthy difference is the abundance of *atalaya* in scientific prose where *centinela* is much rarer. Regardless, the vast majority of *atalaya* and *centinela* tokens belong to higher register texts, with only a small portion of tokens from each term reflecting popular language or speech (1.12% and 3.16%, respectively). Given the small representation of tokens in the popular register for either term, as well as their representation in the nearly identical text categories, we are again reluctant to make any claim as to the preference of speakers for *atalaya* or *centinela* according to register.

In other words, any potential differences in usage based on register are not obvious in the documentation, so we cannot confirm that the factor of register is favorable to the retention of *atalaya*. However, given the general preference for Romance titles of occupations (over Arabism counterparts) in official capacities (e.g., institutions, such as universities and government; cf.

[112] In the seventeenth century, *atalaya* is no longer used in societal prose or legal prose, and instead in journalistic prose.

arquitecto as opposed to *alarife*, *veterinario* over *albéitar*), *centinela* was likely the preferred term for high register use.

4.3.3 Language contact

The final sociolinguistic factor that we will examine in our analysis of *atalaya* and *centinela* is language contact, which will offer valuable information regarding the circumstances of transmission and speaker awareness of the Arabism in question. The reader will recall that, in a broad sense, we categorized our selected Arabisms as direct or indirect borrowings (i.e., transmitted through speech or writing, respectively) and predicted that indirect Arabisms would be more resistant to loss than direct ones. Below we will consider these subfactors of language contact, as well as discover if our data supports our hypothesis of indirect versus direct Arabisms.

4.3.3.1 Transmission

In the previous chapter, we described how contact with Arab agricultural practices and products encouraged the introduction of Arabisms in the domain of agriculture. For the present case study, we are interested in the interaction that occurred through military activity (e.g., border wars, crusades, Reconquest), which required a form of communication that both parties (Arabic and Romance speakers) were able to understand. These situations not only led to instances of individual borrowings but forms of bilingualism as well (Winet 2006: 195).

Given the date of its first documentation in the early eleventh century, *atalaya* entered into Romance during the first period of Moorish invasion and occupation, that is, between the

eighth and eleventh centuries, from the formation of al-Andalus to the peak of its influence. At this time, Winet describes how usage of Arabic by Romance speakers or Romance by Arabic speakers came from a desire to be understood, which gave rise to a form of Arabic in contact (i.e., *árabe en contacto*). The author argues that the cultural prestige of Arabic during this time encouraged some Romance speakers to incorporate Arabisms into their own language (2006: 181).[113]

If under certain circumstances the use of Arabic was due to the desire to communicate in an elementary sense, then in other cases it arose out of necessity. Such is the case of the domains of commerce, politics, and the military, all of which relied, to varying degrees, on bilingual speakers. Over a period of centuries, this Arabic-Romance bilingualism became an indispensable skill for professions in these disciplines. In addition, bilingualism was also important in the eleventh century (and beyond), when royal translators, such as those of Alfonso VI, were essential in the royal courts (Winet 2006: 181).[114]

Many military terms of Arabic origin are first documented in the texts of Alfonso X, where they represent a significant portion of all Arabisms found in said texts. According to García González, many of these terms must have entered into Romance in the early stages of Moorish occupation through the speech of certain kinds of bilinguals, such as soldiers and mercenaries, who had contact with Arabic (1996-1997: 131).

[113] The cultural prestige of Arab culture in the Iberian Peninsula is clear, even centuries after its peak of influence. Consider, for example, the many translations of Arabic texts sponsored by Alfonso X.
[114] For further discussion of the role of translation with regard to Arabisms, refer to section 5.3.3.1 in the following chapter.

Military Arabisms, especially those that refer to terrestrial armed forces, are among those that tend to have an early, popular origin. One example is *alférez* 'second lieutenant; ensign,' whose origin in professional jargon is supported by the restriction undergone by the term as it entered into Romance (cf. Ar. *fāris* 'horseman') (García González 1996-1997:136). Another case is *atalaya*, another early entry into Romance (with a rather large number of terms derived from it) that is documented as early as *Poema de Mio Cid* (c. 1140) and abounds in thirteenth-century texts.[115] As in the previous case of *aceite*, we contend that early documentation and establishment of the term and its derivatives suggest that the transmission of *atalaya* is quite favorable. In other words, the circumstances of contact, especially extended duration of contact, favored the retention of the Arabism.[116] This claim, then, refutes our hypothesis that indirect Arabisms (i.e., those transmitted through writing) were more resistant to elimination than direct Arabisms (i.e., those transmitted through speech).

García González (1993-1994) explains that many scientific Arabisms entered, as one might expect, through writing. That these scientific terms were likely restricted to usage among select bilingual groups only (i.e., the learned elite) disfavored their retention. These same learned terms are also those of which there is the greater awareness of Arabic origin. On the other hand, military terms (and other, quotidian terms), most of which were borrowed through speech, are usually considered to be established Castilian terms, even in the thirteenth century, as demonstrated by their documentation in the works of Alfonso X:

---

[115] García González contends that the appearance of similar terms in other Romance varieties (e.g., Port., Gal. *atalaia*, Cat. *talaia*) helped to establish *atalaya* in Spanish, while its popular origin is supported, like *alférez*, in the significant restriction between the etymon (i.e., *ṭalāyiʕ*, the plural of *ṭalīʕah* 'outpost') and its reflex (i.e., *atalaya* 'diurnal sentinel') in Alfonsine texts (1996-1997: 137).

[116] Thomason (2001) identifies intensity of contact (e.g., duration, relative number of speakers of each language in contact) as one of the social predictors of contact-induced change.

(38) *los cristianos llaman al que este oficio faze alferez*
'the Christians call he who carries out this post alferez'
Siete Partidas, 1491 [a 1265]

(39) *Almocadenes llaman agora a los que antiguamente solian llamar cabdillos de las peonadas*
'What they now call almocadenes they used to call commanders of the troops'
Siete Partidas, 1491 [a 1265]

Naturally, the greater awareness of the Arabic origin of scientific terms left these borrowings susceptible to loss amidst negative attitudes. We will continue our discussion of speaker awareness and attitudes below, addressing how speakers viewed *atalaya* (and military Arabisms in general) neutrally and how neutral attitudes may have protected it (and others) from loss.

4.3.3.2 Speaker awareness and attitudes

Like *aceite*, *atalaya* and its orthographic variants appear to meet the requirements of Castilian phonology well enough that there is little in form alone that would reveal their Arabic origin. For example, there are plenty of Spanish patrimonial terms that end in sequence *-aya* (e.g., *haya* < Lat. FĀGEA, *playa* < Lat. PLAGIA, *raya* < Lat. RADIA, *saya* < Lat. *SAGIA). On the other hand, many Arabisms begin with the sequence *ata-* (e.g., *atabal*, *ataharre*, *ataguía*, *ataifor*, *atanor*, *ataúd*); however, this initial feature could not have been a reliable indicator of origin by any means, since there is also a number of words of diverse origins that begin with the same phonological sequence (e.g., *atar* < Lat. APTARE, *atajar* < *tajar* < Vulg. Lat. TALEĀRE, *atalar* < *talar* < Germanic **tālōn*?, *ataviar*, der. of Goth. **attaujan*, *ataxia* < Gr. *ἀταξία*, etc.).

For these reasons, *atalaya* could not have been identified as an Arabism on a phonological basis alone.

Clearly, in order for speakers to be able to identify *atalaya* as an Arabism, they would have had to consider other features beyond the level of phonology. As far as morphosyntactic features are concerned, *atalaya* is very similar in structure to *aceite*. That is, at the very least, *atalaya* does contain the Arabic article (i.e., *a*(*l*)-), unlike a relatively small number of military Arabisms where it is absent (e.g., *recua* 'train of pack animals,' *zaga* 'rearguard'). Still, while *atalaya* does contain a form of the Arabic article, it is assimilated (and simplified). Of course, Arabisms with assimilated articles (e.g., *adalid*, *arráez* 'commander; skipper; foreman in some jobs') are not as identifiably Arabic in character as Arabisms with unassimilated articles (e.g., *alférez*, *almocadén*), which contain the easily recognizable initial *al-*. Again, with this in mind, initial *a-* is not particularly indicative of the origin of *atalaya*.

Since phonological and morphosyntactic features of *atalaya* are not especially informative with regard to its Arabic origin, speakers might only have been aware of the foreign nature of the word due to the considerable contribution of Arabic to military terminology.[117] Even though *atalaya* does not contain a clear reference to Arabic culture or language like OSp. *mozlemo* 'Muslim' or *mezquita* 'mosque,' its association with Arabic speakers had to have been obvious to many Ibero-Romance speakers during the Moorish occupation of the Peninsula. However, the many military Arabisms that were borrowed initially were largely unknown to monolingual Romance speakers, who encountered them indirectly, through professional jargon

[117] For a more detailed discussion of the contribution of Arabic to Spanish military terminology, we recommend García González (1996-1997).

used by bilinguals, those in direct contact with Arabic, or scholars knowledgeable in Arabic. In addition, any potential association of *atalaya* with Arabic is lost over the centuries; its usage in early modern religious (i.e., Christian) texts as 'guardian, protector' suggests that writers were either unaware of its Arabic origin or that attitudes toward the word were neutral.

Due to its rather inconspicuous form, it is plausible that the typical Castilian speaker, especially after the late medieval period, was unaware of the Arabic origin of *atalaya*. Below we will examine if the documentation reveals any negative attitudes regarding *atalaya*. Since the Spanish elite of the early modern period are generally known for their opposition of Arabisms, we expect that metalinguistic references to *atalaya* may not be favorable to its usage.

Overall, metalinguistic references to *atalaya* are notably neutral. In fact, we encountered no negative remarks on its usage and one evaluation that even appears to be favorable. Of course, neutral usage is not entirely unexpected since the relevant cases here are found in scientific and didactic prose (i.e., dictionaries, especially), which tend to favor objective language, even though there are numerous examples of glaring subjectivity regarding other Arabic terms in prescriptive linguistic texts from the late medieval and modern eras.[118]

(40) *Specula. feminino. es logar donde miran los campos.* ***atalaya****.*
'Specula. feminine. It is the place from which they watch the countryside. atalaya.'
Universal vocabulario en latín y en romance, 1490

[118] One such text is the *Recopilación de algunos nombres arábigos* (1593) by Diego de Guadix, which offers many examples of Arabisms, condemned by the author, as well as Latin or patrimonial substitutes for the nonnative loanwords (e.g., the selection of *muralla* 'defensive wall' over *adarve* 'defensive wall') (Giménez-Eguíbar 2016). Another source of condemnation is *Libro intitulado los problemas de Villalobos* (1543) by Francisco López de Villalobos.

(41) ***Atalaia*** *lugar alto. specula .ae.* ***Atalaia*** *varon que atalaia. speculator.*
'Atalaia high place. specula. ae. Atalaia man who guards. speculator.'
Vocabulario español-latino, 1495

(42) *Episcopus. pi. masculino genero* (...) *especulador o guarda mayor o* ***atalaya***
'Episcopus. pi. masculine gender (...) [equivalent to] especulador or guarda mayor or atalaya'
Vocabulario eclesiástico, 1499

(43) *Speculator. toris. masculino genero* (...) ***atalaya*** *o contemplador o el que esta en logar alto que llaman espacula que contempla della y preuee lo que viene*
'Speculator. toris. masculine gender (...) [equivalent to] atalaya or contemplador or he who is in a high place whom they call espacula, who contemplates from it and forsees what comes'
Vocabulario eclesiástico, 1499

(44) *Lo que agora llamamos* ***centinela****, amigos de vocablos extranjeros, llamavan nuestros españoles: de noche, escucha, y de dia,* ***atalaya****, nombres harto más propios para su oficio*
'What we now call centinela, friends of foreign words, our Spaniards called: at night, escucha, and by day, atalaya, names that are much more suitable for the profession'
De la guerra de Granada, 1569-1573

(45) *+****ATALAIA****. (Añade). Latine specula, ae [en latín, specula, ae,* ***atalaya****] y del sonido de este nombre engañados, los Idiotas inventaron que en algunas de estas torres avía çiertos espejos, en los quales se representaban los navíos que venían por la mar por muy gran distançia, a la qual no alcançaba la vista de los que atalayaban en las dichas torres.*
'+**ATALAIA**. (Add). Latin specula, ae [in Latin, specula, ae, **atalaya**] and deceived by the sound of this name, the idiots invented that in some of these towers there was some kind of mirror, in which were represented the ships that were coming by sea from a great distance, to which the view of those on guard in said towers did not reach'
Suplemento al Tesoro de la lengua española castellana, c. 1611

In addition to their overall neutrality, Diego Hurtado de Mendoza provides an example in which the Arabism is even favored, suggesting that *atalaya* (and *escucha*) may have been

preferred to its Italianism counterpart *centinela* at the time. Furthermore, in these examples, *atalaya* never appears with commentary on its origin, despite being found alongside Latin equivalents. Consider another example, a passage that describes the military structure and posts of Christian forces (adopted from their Arabic-speaking counterparts):

(46) *Onrradamente estableçieron los antiguos que fuesen fechos los* ***adalides*** *segund en la ley* (...) *E ellos han poder de mandar a los* ***almogauares*** (…) *and a los peones and de poner de dia* ***atalayas*** *and de noche escuchas and* ***rodas*** *and han de ordenar las* ***algaras*** (...) *and ellos han poder de fazer* ***almocadenes*** *a los peones* (...)
'Honorably, the ancients established that they were made the leaders according to the law (…) And they have the power to order the mercenary raiders (…), and the peons, and post sentinels during the day, and sentinels at night, and patrols, and they shall order the raids (…), and they have the power to make troop captains of the peons (…)'
Siete Partidas, 1491 [a 1265]

For these reasons, it is possible that speakers were generally unaware that *atalaya* and other military Arabisms were words of Arabic origin. If this is accurate, then speaker awareness and attitudes together may be considered a factor that is favorable to the retention of *atalaya* since it appears that the Arabism was not condemned as others had been. On the other hand, if speakers did recognize *atalaya* as an Arabism, generally neutral attitudes toward military Arabisms may explain their resistance to loss (García González 1993-1994: 354, 1996-1997: 138).

Elsewhere in the broader category of military posts, attitudes are generally neutral as well. On one hand, *alférez*, *adalid*, and other such terms do not appear to have experienced the semantic depreciation that other trades had; for example, *alférez* continues to be used as an official title in the Spanish armed forces. This lack of semantic pejoration contrasts with other positions or trades, which fall outside of the military domain (e.g., *alarife* 'architect' >

'construction worker; carpenter,' *alcahuete* 'mediator in marriages' > 'pimp, procurer) (García González 1993-1994).

4.4 Summary

In this chapter, we have addressed a number of linguistic and sociolinguistic factors that have had an impact on the retention of *atalaya* and other military Arabisms. Among the most relevant of these factors, as in the previous chapter, are semantic differences and language contact (i.e., transmission, speaker awareness and attitudes). Additionally, we consider the success of related vocabulary to be a factor of secondary importance in terms of retention, while formal variation is favorable as well. While the role of register in the retention of *atalaya* is inconclusive, dialect does not appear to play a role at all.

The introduction of *atalaya* may have been motivated by the low frequency of the reflexes of Lat. SPECULA 'watchtower' and VIGIL 'sentinel' in Hispano-Romance. Without well-established Latin equivalents, Romance speakers borrowed *atalaya* to represent 'sentinel' and 'watchtower' in Hispano-Romance, in a semantic category that was evolving at the time of Moorish invasion and occupation, when Romance speakers were exposed to new military concepts and terminology. Weinreich (1968) contends that low frequency is an internal factor of languages that makes words more susceptible to replacement, using contact between Russian and Finnish as an example, explaining that the infrequent use (and variation) of names of parts of tools in dialectal Russian has led local varieties to adopt equivalent Finnish terms in bilingual areas.

First recorded with the metonymic meaning of '(diurnal) sentinel,' usage shows how writers also used *atalaya* with the prototypical meaning of 'watchtower,' a stable meaning that the term retained. Once the term was well established, *atalaya* further broadened in meaning through regular usage, so that the diurnal feature of 'diurnal sentinel' became increasingly irrelevant; additionally, through metaphor, *atalaya* began to be used as 'guardian.' The introduction of its Italianism counterpart, *centinela*, occurs during the lexical transition of Castilian, during which writers actively selected borrowings from Latin and Romance languages that were more semantically precise than patrimonial counterparts. After all, *atalaya* for centuries had been polysemic and, therefore, ambiguous. *Centinela* is first introduced with the meaning of '(nocturnal) sentinel' and experiences a semantic generalization similar to that of *atalaya* (> 'sentinel' > 'guardian; sentinel'). Once *centinela* was established and as broad as its counterpart, the terms must have been very close in meaning. With the coexistence of two synonyms and the ambiguity of *atalaya*, the 'sentinel' meaning of *atalaya* begins its descent into obsoletion, while still retaining its meaning of 'watchtower.' The introduction of another, less ambiguous word[119] for 'sentinel' (i.e., *centinela*) certainly seems the most plausible linguistic factor for the replacement of *atalaya* as 'sentinel.' Furthermore, the loss of the 'sentinel' (rather than the 'watchtower') meaning, with a [+ human] feature, may have been aided by generally negative attitudes associated with Arab individuals, a feature absent from the 'watchtower' meaning that was maintained.

Muñoz Núñez (1999) explains how the phenomenon of polysemy leads to ambiguity and how languages evolve to amend it. Some of the ways in which languages have evolved to

[119] Again, note the potential for ambiguity in *atalaya*, which for centuries could mean both 'sentinel' and 'watchtower.'

eliminate the ambiguity resulting from polysemic words include: (1) differentiation in grammatical gender (Sp. *el cura* 'priest' vs *la cura* 'cure'), (2) differentiation in the formation of plural forms (Fr. *aïeuls* 'grandparents' vs *aïeux* 'ancestors'), (3) differentiation in word order (Sp. *un hombre pobre* 'an indigent man' vs *un pobre hombre* 'an unfortunate man'), (4) additions to the word in question (OSp. *atalaya* 'watchtower; sentinel', *torre atalaya* 'watchtower'), (5) differentiation in form (Eng. *discreet* 'having good judgement; unnoticeable' vs *discrete* 'individually distinct; noncontinuous'). When these linguistic resources fail, one or more meanings of a given word are generally lost.[120] Rudskoger (1952) examined some 120 English adjectives and discovered that only three of the many polysemic adjectives studied had been eliminated and replaced. Regardless of the cause of decline of individual words, Muñoz Núñez (1999) claims that more often than not, words that demonstrate polysemy are not lost entirely, but rather they lose one or more of their meanings. This process may explain the evolution of *atalaya*, in which one of its meanings (i.e., 'sentinel') is lost, another term (i.e., *centinela*) takes its place, and the other meaning of *atalaya* (i.e., 'watchtower') is retained.

The other significant consideration for the retention of *atalaya* is language contact, that is, how transmission and speaker awareness and attitudes affected its survival. Like many military Arabisms, *atalaya* is popular in origin. Although we hypothesized that Arabisms of popular origin (i.e., direct borrowings) were generally more susceptible to loss than those of learned origin (i.e., indirect borrowings), military Arabisms may be more resistant due to the

[120] Although there are hints of differentiation in grammatical gender (e.g., *el ~ la atalaya* 'sentinel,' *la atalaya* 'watchtower') and additions to the word in question (e.g., *torre de atalaya* 'watchtower'), these strategies are never generalized before the elimination of the 'sentinel' meaning.

intensity (i.e., duration) of military conflict and, therefore, language contact between Romance and Arabic speakers during the Reconquest.

As far as awareness and attitudes are concerned, *atalaya* divulges little of its Arabic origin in terms of its phonology, morphosyntax, or semantics. In other words, based on appearance alone, *atalaya* easily fits alongside other Spanish terms of similar formal features. The same appears to be true of its semantics since its meaning does not have any overt reference to Arabic culture or language. This lack of overtly Arabic character may be the main reason for the absence of subjective commentary regarding the term in the documentation. If *atalaya* was indeed undetected as an Arabic loanword, its lack of noticeable features likely favored its retention since Arabisms were erased on such a large scale, simply for their origin.

Finally, we will consider how *atalaya* behaves in relation to the other military Arabisms introduced into Castilian Romance. Formal variation of *atalaya* is rather limited compared to other similar terms of the first wave of military Arabisms. Consider, for example, the considerable variation of *alcaide*, *alférez*, and *anúbada*.[121] However, *atalaya* is not frequently documented until the thirteenth century, and in this way, *atalaya* more closely resembles the second wave of military Arabisms, which generally demonstrate little formal variation: *adarga*, *arriaz*, *alfanje*.[122] The development of the semantic category in question is more complicated, with both waves containing some military Arabisms that have not generated additional terminology; however, the fact that *atalaya*, *algara* (also introduced in the first wave but not well

[121] *Alcaide* is documented with more than a dozen formal variants, *alférez* some two dozen, and *anúbada* some three dozen (Corriente 2008: s.v. *alcaide*, *alferes*, *anúbada*).
[122] Corriente identifies three formal variants of *adarga*, two of *arriaz*, and one of *alfanje* in Castilian (2008: s.v. *adarga*, *arriaz/l*, *alfanje*).

established until the second) and *rebato* (part of the second wave) are among the most productive lexical bases suggests the greater potential for derivation of military terms that were established in the second wave. Like many other military Arabisms, *atalaya* is a (partial) synonym of other Romance terms (cf. *alcázar* 'fortress; palace' vs *fortaleza* 'fortress,' *adalid* 'leader; guide' vs *caudillo* 'leader; dictator.' However, most retained military Arabisms also contain nuances in meaning that the Romance counterparts do not, which may favor their retention: *adarga* 'leather shield' vs *escudo* 'shield,' *alfanje* 'Moorish saber' vs *sable* 'saber.'

As we indicated above, *atalaya*, like other retained military Arabisms, is generalized in its geographic distribution. In cases of competing Romance synonyms (e.g., *zaga* 'rearguard' vs *retaguardia* 'rearguard'), we contend that other military Arabisms, like *atalaya*, are likely preferred in popular language, whereas the Romance counterpart is preferred in formal language (with the understanding that our data were unable to confirm this preference for *atalaya*). With regard to language contact, the transmission of *atalaya* occurs during the first wave of military Arabisms, but given its limited formal variation and large capacity for derivation it behaves more like a military Arabism in the second wave, many of which were retained, often with subtle differences in meaning from Romance counterparts. Also, as in the case of many military Arabisms from the second wave, *atalaya* is viewed neutrally. With these factors in mind, we argue that *atalaya* behaves in a manner similar to many other military Arabisms, especially those introduced in the second wave. Furthermore, we believe that other studies may further corroborate our results for military Arabisms upon closer inspection of individual borrowings.

5 Case study: *jaqueca, hemicránea* and *migraña*[123]

In our final case study, we will explore one more Arabic-Romance lexical group, that of *jaqueca*, *hemicránea*, and *migraña*, three synonyms that we have chosen to represent the wider domain of medicine, remedies, and illnesses. Through an analysis of several linguistic and sociolinguistic factors, along with the historical context in which the sciences and medical knowledge became widely available in medieval Iberia, we will determine which of these factors are favorable, unfavorable, or irrelevant to the maintenance of the Arabism in question (i.e., *jaqueca*). We contend that *jaqueca* was introduced in the absence of an equivalent patrimonial term for the concept of 'migraine' and that its retention is the result of several favorable factors, most notably limited formal variation, distinctive register usage, and neutral attitudes toward the term. Through an examination of this lexical group and others, we will gain a better understanding of the conditions under which particular medical Arabisms were retained in spite of generally negative attitudes toward Arabic and the emergence of synonymous Romance terms.

In section 5.1 below, we provide an overview of the transmission of science and medical knowledge in medieval Iberia, focusing on the contributions of the Arabs (and other Arabic-speaking individuals) to the field of medicine. In section 5.2, we consider the influence of linguistic factors (i.e., formal variation, related vocabulary, semantic differences) on the preservation of *jaqueca*. In section 5.3, we examine sociolinguistic factors (i.e., dialect, register, language contact) with that same objective. Through our evaluation of factors, we will once more

[123] Herrera (1996) and Herrera, Sánchez González de Herrero and Zabía (1997) place the various medieval forms of this term (e.g., *emigranea*, *migranea*) under the heading of *hemicránea*, a learned form (s.v. *hemicránea*). The Latinism *hemicránea* differs from *migraña*, identified by Covarrubias as a Valencian borrowing (1611: *axaqveca*).

rely on written documentation from CORDE (formal variation, semantic differences, register), historical and contemporary dictionaries (related vocabulary), and lexical studies (dialect, language contact) to explain the role that each has had on the Arabism. In the last section of the chapter, 5.4, we compare the treatment of *jaqueca* as a medical Arabism and demonstrate how it differs from other such terms.

5.1 The sciences and medical knowledge in medieval Iberia

From the mid-eighth century to the beginning of the thirteenth, the Arabs were responsible for the spread of culture and civilization throughout much of the globe, including (and especially) the Iberian Peninsula. Through these individuals, texts and knowledge from Antiquity passed onto medieval Western Europe, from geography to astronomy and mathematics, to philosophy, botany and medicine (Hitti 1951: 557-590).[124] Acknowledging the foreign (e.g., Latin, Persian, Greek) origin of these natural sciences[125] and others, Muslim scholars transmitted knowledge of these disciplines primarily through translations from Arabic to Latin, a practice that originated in the eastern Islamic empire (Chejne 1974: 344-345).

Prior to the translation of the natural science texts (e.g., ninth through eleventh century), intellectual efforts were instead largely devoted to religion or to Arabic studies (e.g., grammar, philology). As a culturally conservative territory, al-Andalus was initially resistant to the translation of texts from foreign cultures; however, as translations of ancient texts on natural

[124] These are only some of the more notable areas of contribution of the Arab peoples to the Iberian Peninsula (and Western Europe, in a broader sense), but they are by no means exhaustive. Furthermore, for the purposes of this chapter, we will focus primarily on the contribution of medical knowledge and terminology.
[125] Chejne (1974: 344) clarifies that the Muslim conception of a natural science is one that is "natural to man and not restricted to any particular religious group." Under this definition, the following are considered natural sciences: philosophy and wisdom (e.g., logic, physics, metaphysics, medicine) and mathematics (e.g., geometry, arithmetic, music, astronomy).

sciences proliferated in the Islamic East, so did interest grow in al-Andalus. The earliest sciences to enter into the culture of al-Andalus were the most practical and useful for quotidian life, that is, astronomy, mathematics, and, of course, medicine (Chejne 1974: 345).

Having inherited the medical knowledge of Greece and the East, the Arabs advanced the discipline in their own right. Translating from Greek to Arabic, they were introduced to Hippocrates and Galen, the best known and most highly regarded of Greek physicians in the Arab Empire. As medicine secured its position as a science (as opposed to superstition or witchcraft in much of medieval Europe), so did the establishment of the institution of hospitals, which were constructed in the principal cities of the Islamic Empire, including Baghdad and Cordova. While much medical knowledge was acquired from other civilizations, the Arabs played an important role in its preservation and transmission to al-Andalus, as well as through their own findings (Chejne 1974: 350-352).

The most important of Islamic physicians, al-Razi (865-925) composed some 200 works, addressing various fields of natural sciences (e.g., astronomy, chemistry, medicine, etc.). His seminal work, *al-Hawi*, contains medical knowledge from his own observations, along with that of past physicians. In 1279, *al-Hawi* was translated into Latin, and as such was available for use in European universities. Another noteworthy scholar is Ibn Sina (980-1037), whose *al-Qanun fi-l-tibb*, a comprehensive work containing much of the medical knowledge available in his time, was translated into Latin in the twelfth century (Chejne 1974: 352-355). Abulcasis (936-1013), a physician and surgeon, is best known for *al-Tasrif*, a work mostly based on the medical knowledge of his predecessor, al-Razi, but it also contains contributions regarding surgery and

surgical instruments (Hitti 1951: 576-577). This text was also translated into Latin in the twelfth century.

Flourishing especially in the tenth and eleventh centuries (Moreno Fernández 2005: 68), the production of Arabic language medical texts was substantial, at a time when Europeans still relied on superstitions. However, through contact during the Arab occupation of the Iberian Peninsula and elsewhere, Christians would become receptive to the medical contributions of the Arabs (Chejne 1974: 357-358). In order to disseminate this new knowledge, many Arabic texts were translated into Latin, especially in the twelfth and thirteenth centuries. At this time, a school of translators was established in Toledo to pass on the knowledge of various scientific disciplines from Greek, Latin, and Arabic texts to Romance. The translations composed by this school were often collaborative: learned in Latin/Romance and Arabic, Jewish scholars would translate Arabic works into Romance, while Christians would prepare the Latin translations (Read 1974: 174-175).

As the field of medicine grew in the Middle Ages, so did the terminology that accompanied it. Through the translation of Arabic medical texts in al-Andalus, many medical Arabisms entered into Hispano-Romance. Introduction of foreign medical terms must have been favorable in cases where there was no lexical equivalent, or if the patrimonial term had fallen into disuse (cf. McMahon 1995: 201), the first of which appears to be the case of *jaqueca*, whose Latinism equivalent (i.e., popular forms of *hemicránea*; e.g., *migranea*) is only sporadically represented in the documentation. *Jaqueca*, on the other hand, is far more successful, having strong representation in high register medical texts; however, *jaqueca* must also have been

particularly well received and established in speech, given its documented phonological forms (e.g., OSp. *xaqueca* /ʃakéka/, rather than a Latinate form such as **saqueca*).[126] Such circumstances, namely, a secured position in speech communities and the preceding need to fill a lexical void (i.e., a necessary, or cultural, borrowing), are very favorable to the maintenance of this Arabism.[127] In the sections that follow, we will examine other favorable factors in the retention of *jaqueca*, including some that are unfavorable or irrelevant.

5.2 Linguistic factors

Now that we have addressed the historical context in which much medical knowledge was transmitted (section 5.1), we may discuss the relevance of linguistic factors in the retention of *jaqueca* despite the introduction of a synonymous Romance term. We will first discuss formal variation (section 5.2.1), followed by related vocabulary (section 5.2.2), and lastly semantic differences (section 5.2.3). In general, restricted variation in form, a well-established group of related vocabulary, and observable semantic differences are favorable to the retention of Arabisms. In the following sections, we will see that not all linguistic factors are favorable to the maintenance of *jaqueca*; however, it is not necessary that all linguistic factors favor a given Arabism for it to be maintained.

5.2.1 Formal variation

Given that variation in form is unfavorable to the retention of a loanword (Cano Aguilar 1993, Dworkin 1989, Álvarez de Miranda 2009), an Arabism of relatively little formal variation

[126] For further discussion of the transmission of *jaqueca*, see the sections on register (5.3.2) and language contact (5.3.5).
[127] Refer to section 2.3.1 for a definition of a necessary (or cultural) borrowing.

should be more resistant to loss than one of greater variation. With this in mind, let us consider the orthographic variants of the Arabism in question, *jaqueca*, and its Latinate counterpart, *hemicránea*. Even a cursory examination of these terms demonstrates the considerable number of variants that have been documented in this group of words, as we will explain below.

Table 5.1. Distribution of *jaqueca* variants among selected writers, 1250-1614[128]

Author	Variants	Number of tokens	Representation (%)
Anonymous$_1$	*axaqueca*	2	66.67
(a 1500)	*enugeça*	1	33.33
Anonymous$_2$	*ajaqueca*	1	50.00
(1600-1604)	*jaqueca*	1	50.00
Góngora y Argote	*ajaqueca*	1	50.00
(1580-a 1627)	*xaqueca*	1	50.00
Gómez Miedes	*jaqueca*	1	50.00
(1589)	*xaqueca*	1	50.00
Enríquez	*axaqueca*	2	50.00
(1471)	*xaqueca*	2	50.00
Méndez Nieto	*axaqueca*	3	50.00
(1606-1611)	*xaqueca*	3	50.00

As in the case of *atalaya* in the previous chapter, no writer in CORDE uses more than two orthographic variants of *jaqueca*. Where they differ, however, is the number of authors who use two variants of *jaqueca*. Although the number of tokens is admittedly small, Table 5.1 suggests that authors do not have a preference for any particular formal variant of *jaqueca*: In the time period studied, five of the six authors who employ two variants use both in equal measure, whereas only one author uses *axaqueca* (66.67%) more than the peculiar variant that is *enugeça*

[128] Although we examined both singular and plural forms of *jaqueca*, the selected writers only employed singular ones. Furthermore, Luis de Góngora y Argote is a source of *jaqueca* tokens dated 1580-a 1627; most of this date range falls within the time period studied.

(33.33%).[129] Again, the sample size is too small to draw conclusions from a single author, but the overall impression indicates the persistence of multiple *jaqueca* variants well into the early modern era.

Let us now consider the formal variants of *jaqueca* in their entirety. Although there are two cases of *alxaqueca*,[130] this variant is rare, appearing only in a single text (i.e., *La Lozana Andaluza*). The same is true of *enugeça*, which is documented only once, in a text known as *Tratado de patología* (CORDE) and *Tratado medico* (Herrera, Sánchez González de Herrero and Zabía 1997).[131] The remaining, and much more frequent, orthographic variants may be divided into two groups: those with the assimilated Arabic article (i.e., *ajaqueca*, *axaqueca*) and those without it (i.e., *jaqueca*, *xaqueca*).

Table 5.2. Orthographic variants of *jaqueca* in all documents from Spain, c. 1250-1614[132]

Variant	**First documentation**	**Text**
xaqueca	c. 1250[133]	*Lapidario*
axaqueca	1438	*Arcipreste de Talavera* (*Corbacho*)
enugeça	a 1500	*Libro de recetas*
jaqueca	1521-1543	*Epístolas familiares*
alxaqueca[134]	1528	*La Lozana Andaluza*
ajaqueca	1580-a 1627	*Romances*

[129] Due to the discrepancy between orthographic variant *enugeça*, on one hand, and the original source (i.e., Hispano-Ar. *šaqíqa*) and all other variants, on the other, we have to imagine the possibility that *enugeça* is a copy error.
[130] We encountered two cases of *aljaqueca*, but not *alxaqueca*, in CORDE, despite the fact that Corominas and Pascual (1980-1991: 492) note the appearance of the latter in *La Lozana Andaluza*, which appears in said corpus. This discrepancy is due to the modernization of the original spelling (i.e., *alxaqueca*) in this 1994 edition of the text.
[131] To avoid potential confusion, we are referring to the text that the Hispanic Seminary of Medieval Studies identifies as TRM.
[132] See note 6.
[133] The first documentation of *jaqueca* (as *xaqueca*) is certainly in the *Lapidario* of Alfonso X. However, there is debate whether this case is simply a use of a foreign term that had not yet been incorporated into Spanish. We will briefly resume our discussion of this matter in section 5.3.3.1 below.
[134] See note 8.

With *enugeça* as an exception, orthographic variants of *jaqueca* are relatively invariable, especially given the tendency for errors to arise from the transliteration from Arabic to Latin script (Herrera and Vázquez de Benito 1985: 74). Pensado Figueiras (2012: 399) states that the manuscript of the *Libro de recetas* shows that *enugeça* was written with a steady hand but suggests that this variant may have been influenced by the gloss that follows it (i.e., *que es dicha enfermedat de la cabeça*).[135] However, the rest of the form (i.e., *enug-*) is uncertain. Herrera (1996), on the other hand, includes *enugeça* under the *jaqueca* entry but offers no explanation of the unusual variant. Still, the existence of a wide range of formal variation is not uncommon for medical terms of Arabic origin, such as Ar. *aṭṭāʕūn* 'plague, epidemic': *althaun, althohoin, althoin, alchoboin, taon, thabun, thahaum* (Corriente 2008: s.v. *althaun*). Such formal variation in medical Arabisms is the result of several factors described by Herrera and Vázquez de Benito (1985: 74): (1) the lack of familiarity of scribes with medical language, (2) the absence of equivalent Latin medical terminology, leading to diverse transliterations, for which there was no established norm, and (3) mistakes that occur from a general lack of familiarity with the material. Another important factor to consider is the manner of transmission (i.e., written vs oral), as it is possible for borrowings to experience greater variation through speech.

Like *jaqueca*, the formal variants of *hemicránea* are numerous, but given its transmission from Latin the latter term is surprisingly even more variable than its Arabic-based counterpart. Consider, for example, the orthographic variation present in *Tratado de patología* alone:

Table 5.3. Distribution of *hemicránea* variants in *Tratado de patología* (TRM) (a 1500)

[135] Note the ending (*-eça*) that *enugeça* and *cabeça* share.

Variant	Number of tokens	Representation (%)
emigranea	1	12.50
emjgranea	1	12.50
enemjgranea	1	12.50
engranea	1	12.50
enjmjclanea	2	25.00
enjmjgranja	1	12.50
njgramja	1	12.50

This text demonstrates highly variable orthography with regard to the lexical item in question, providing more formal variants of *hemicránea* (n = 7) in a single text than the total number of variants in all documentation of *jaqueca* (n = 6) during the time period studied. Such variation of *hemicránea* suggests that author had little or no familiarity with the term and perhaps, more generally, the material (i.e., medicine). In fact, few of the various transliterations in *Tratado de patología* are alike, other than *emigranea* and *emjgranea*, both of which are equivalent to /emigɾánea/ or /emigɾánia/. In the variant *engranea*, the absence of *i* represents what appears to be syncope of the unstressed vowel /i/, alongside a substitution of /m/ for /n/, while three others (i.e., *enemjgranea*, *enjmjclanea*, *enjmjgranja*) contain non-etymological initial *en-*, with *enjmjclanea* even containing what appears to be a consonant sequence /kl/ in place of /gɾ/.[136] Finally, variant *njgramja* demonstrates what amounts to a combination of apheresis of its initial vowel and metathesis of its nasal consonants.

Finishing out the remainder of the formal variants, in other texts, are the following diverse forms: *emicranea*, *evigrançe*, *migranea*, *migraña*, *migraneum*, *mjgraneam*, and *milgrania*. Of these medieval variants, *emicranea* most closely resembles its etymon (Late Lat.

[136] Sánchez González de Herrero (1998) identifies both western (e.g., confusion of *l* and *r*) and eastern linguistic features (e.g., plural forms in *-es* for *-as*) in the text, arguing that it is Aragonese in origin but that its copyist was from western Iberia (1998).

HEMICRANIA). *Evigrançe*, on the other hand, is unusual: Herrera proposes that the form *evigrançe* in *Libro de recetas* is a potential error for *enigranea* (1996: s.v. *hemicránea*), whereas Pensado Figueiras proposes that *evigrançe* is a deformation of *emigranea* (2012: 399).[137] Variants *migranea* and *mjgraneam* are similar, both demonstrating apheresis of initial vowel /e/ and representing either /migɾánea/ or /migɾánia/. *Migraña* appears to contain palatal /ɲ/ instead of the more common sequence of /ne/ or /ni/. The final *m* of *mjgraneam*, in *Cirugia*, may be evidence of the learned nature of the term, with the author making an effort to reflect this in its Latinate appearance. Variant *migraneum* is similar in this regard, although the final *a* has been replaced with *u*, while the author of *Tractado contra el mal serpentino* inserted non-etymological *l* in *milgrania*. A summary of the formal variation of *hemicránea* is found in Table 5.4 below.

Table 5.4. Orthographic variants of *hemicránea* in all documents from Spain, c. 1400-1614[138]

Variant	First documentation	Text
migranea	c. 1250	*Lapidario*
emicranea	1254-1260	*Judizios de las estrellas*
mjgraneam	1440-1460	*Cirugia* (TED)
emigranea	1494	*Compendio de la humana salud* (CHS)
migraneum	1494	*Compendio de la humana salud* (CHS)
migraña	1494	*Traducción de El Libro de Propietatibus Rerum*
evigrançe	a 1500	*Libro de recetas* (RES)
emjgranea	a 1500	*Tratado de patología* (TRM)
enemjgranea	a 1500	*Tratado de patología* (TRM)
engranea	a 1500	*Tratado de patología* (TRM)
enjmjclanea	a 1500	*Tratado de patología* (TRM)
enjmjgranja	a 1500	*Tratado de patología* (TRM)
njgramja	a 1500	*Tratado de patología* (TRM)

[137] This is a reasonable claim: Even given the diversity of orthographic variation of *hemicránea*, *evigrançe* is notably different from the other formal variants, uniquely featuring consonants *v* and *ç*.
[138] This table synthesizes data from CORDE and Herrera (1996). Note that the modern, learned form of *hemicránea* does not surface in CORDE until centuries later, in 1807.

milgrania	1542	*Tractado contra el mal serpentino* (YSL)

Comparing the above table with Table 5.2, it is clear that *jaqueca* has the advantage over *hemicránea* in the way of formal variation in two respects: not only does the Arabism have fewer formal variants overall than its Romance counterpart, but *jaqueca* variants are also more similar (in orthography and phonology) to one another than those of *hemicránea*. Herrera and Vázquez de Benito claim that *jaqueca* had already been in widespread use well before its proliferation in the written register (1985: 99); in this case, the greater familiarity of speakers (and writers) with the Arabism than with *hemicránea* explains its relatively consistent formal variation. In short, by presenting fewer, less varied variants than its patrimonial synonym, the factor of formal variation is certainly favorable to the retention of *jaqueca*.

Other medical Arabisms that have demonstrated significant formal variation are typically lost, as in the case of *aprea* and *botor*. *Botor* is documented with at least six variants, most of which are considerably different in form from one another: *abrojos*, *abusos*/*abuzos*, *batahara*, *bothor*/*botor*. *Aprea* is likewise variable, documented with at least five variants: *alabirati*, *alhebria*/*alhebrie*, *aprea*/*area*. The maintained, and current, medical Arabisms are generally far less variable in form (e.g., *aliacán*, *zaratán*). The variants of *aliacán*, while documented as four forms, roughly share the shape, that is, /aliak(ɾ)a(n)/. *Zaratán* is further restricted in form, documented only as *çaratán* and *zaratán*.[139]

5.2.2 Derived and related vocabulary

[139] See Corriente (2008: s.v. *alacrán*, *alhebria*, *botor*, *saratá*).

Words derived from *jaqueca*, *hemicránea*, and *migraña* are fewer than those derived from *aceite* / *olio* / *óleo* or even *atalaya* / *centinela*. In fact, they are so few that we have found it appropriate to extend our scope to other, related words and those derived from them to better understand the behavior of the semantic category in question, namely, the head and its ailments and conditions.

As we have already discussed, the words that have been documented with the meaning of 'migraine' in the Middle Ages are *jaqueca* (c. 1250 *xaqueca*) and *hemicránea* (c. 1250 *migranea*). In addition to these, we will include *migraña*, which shares an etymon (i.e., Late Lat. HEMICRANIA) with *hemicránea*, but it is considered a separate term. What is unusual about this doublet is that there is no significant denotative meaning between *hemicránea* and *migraña*.[140] Of these, there is no written record of *hemicránea* having given rise to any derived words, while *migraña* has given *migrañoso* (1966) 'relative to migraine; suffering from migraine.' The Arabism *jaqueca* has produced *ajaquecado* 'suffering from migraine,' *ajaquecar* '(to) give or cause migraine,' *ajaquecarse* (1580-a 1627) '(to) suffer from migraine,' and *jaquecoso* (1878) 'causing migraine; annoying, bothersome, irritating.'[141] What *migrañoso*, *jaquecoso*, and *ajaquecado* have in common is that they are all attributes. Although these three terms are all documented as literal adjectival equivalents of their respective sources (i.e., noun or verb), *jaquecoso* has also undergone a metaphorical extension of meaning, with the conceptual link between migraine and irritation being its now obsolete meaning of 'causing migraine.' This

[140] Usually, etymological doublets have different meanings: *causa* 'cause' / *cosa* 'thing,' *computar* 'compute' / *contar* 'count.' In this case, *hemicránea* and *migraña* are defined identically as medical terms for *jaqueca* 'migraine.'

[141] In addition, Garulo indicates the existence of a local expression, *ponerse jaquecoso*, '(to) half cry, half groan' in one site in Huelva (1983: 116).

same link is also demonstrated in the other two terms derived from *jaqueca*, namely, *ajaquecar* and *ajaquecarse*, which are also in disuse.

Like 'migraine,' there are also three words with the broader meaning of 'headache': *dolor de cabeza* (1254-1260 *dolor de cabeça*), *cefalea* (a 1450 *çefalea*), and *cefalalgia* (1555). First documented in *Judizios de las estrellas*, *dolor de cabeza* is the quotidian (and only) term for 'headache' for more than a century. Later, learned *cefalea* (< Late Lat. CEPHALAEA) and *cefalalgia* (< Lat. CEPHALALGIA) are introduced as specialized terms in the domain of medicine (cf. Eng. *cephalgia*, *cephalalgia*). Although there are no terms derived from *dolor de cabeza* (already a compound term) or *cefalea*, *cefalalgia* has given rise to *cefalálgico* 'relative to headache,' which is also an adjectival equivalent of its source (i.e., noun) and comparable to the 'relative to migraine' meaning of *migrañoso*.

Other terms that refer to conditions related to the head are those derived from the base *cefal*(*o*)- (< Gr. *-képhalos,* from *kephalḗ* 'cabeza'), which in most cases is combined with the suffix *-ia* but also with *-itis*. Some of the many such lexical items include (but are not limited to) *acefalia* 'acephalia, absence of a head,' *bicefalia* 'presence of two heads,' *braquicefalia* 'brachycephaly,' *cefalitis* 'inflammation of the head,' *dolicocefalia* 'dolichocephaly,' *encefalitis* 'encephalitis,' *hidrocefalia* 'hydrocephalus,' *mesocefalia* 'mesocephaly,' and *tricefalia* 'presence of three heads,' as well as the corresponding adjectives of the nouns ending in *-cefalia* (i.e., *acéfalo, bicéfalo, braquicéfalo, dolicocéfalo, hidrocéfalo, mesocéfalo, tricéfalo*). As the reader will no doubt realize, these specialized words primarily lend themselves to the domain of medicine (although *bicefalia*, *tricefalia*, and their adjectives may be used metaphorically and

therefore outside of this domain). Having Greek or hybrid Latin-Greek etymology, all of these terms are learned and belong to the formal register.

Through this brief examination of terminology in the semantic category of the head and its ailments and conditions, it is clear that there is a division between the quotidian (i.e., *dolor de cabeza*, *jaqueca* and its derivatives) and specialized (i.e., derivatives of *migraña*, *cefalalgia*, and especially the base *cefal*(*o*)-) terminology that recalls that of *aceite* and *óleo*, whereby the learned terms are greater in number, but those of popular speech are greater in vitality. This, of course, is in large part due to the combinatory potential of the broad meaning of *cefal*(*o*)- for the domain of medicine. On a smaller scale, in the realm of words based on the meaning 'headache' or 'migraine,' it is the Arabism that has been the most productive lexical item, rather than any of the Romance or learned terms. In other words, of the six contemporary terms meaning 'migraine' (*jaqueca*, *hemicránea*, *migraña*) or 'headache' (i.e., *dolor de cabeza*, *cefalea*, *cefalalgia*), *jaqueca* has given four derived words, in contrast to the one derived word each for *migraña* and *cefalalgia* or the absence of such words derived from the remaining terms. In all, the greater derivative output of *jaqueca* suggests that the Arabism is among the most established of these terms in this semantic category. As such, the creation of a modest number of lexical items related to *jaqueca* is favorable to its retention.

Of the medical Arabisms addressed by Herrera and Vázquez de Benito (1981, 1982, 1983, 1985), we found that only three other such terms gave rise to additional terminology: *aliacán* 'jaundice' and *aliacanado* 'jaundiced,' *botor* 'bubo' and *botoral* 'relative to a bubo; similar to a bubo,' *hadruba* 'hump' and *hadrubado* 'hunchbacked.' *Safena* and *zaratán*, on the

other hand, although in use today, have not served as the base for other words. Viewed in their entirety, then, the Arabisms that the authors examined are not nearly as productive as *jaqueca*, with a total of four derived lexical items. If these examples are indicative of medical Arabisms more generally (with most giving rise to one or no derived words), *jaqueca* is an unusually productive and, therefore, well established medical Arabism.

5.2.3 Semantics

As in the other lexical groups we studied, the words for 'migraine' in the Romance languages fall into two broad categories: (1) Arabisms and (2) patrimonial terms or Latinisms. On one hand, only a few Romance languages have an Arabic borrowing for 'migraine,' as demonstrated in Port. *enxaqueca*, Gal. *xaqueca*, Sp. *jaqueca* (< Hispano-Ar. (*iš*)*šaqíqa*). On the other hand, most Romance varieties contain patrimonial terms or Latinisms for the concept, as in Port., Gal. *hemicrania*, Sp. *hemicránea*, Cat. *hemicrània*, Fr. *hémicrânie*, It. *emicrania* (< Late Lat. HEMICRANIA < Gr. *hēmikranía*). Some of these languages even contain an additional term for 'migraine,' as in Sp. *migraña*, Cat. *migranya*, Fr. *migraine*.

A patrimonial term for 'migraine' may never have existed given the absence of such a term from the documentation. Furthermore, several passages from the *Lapidario* by Alfonso X suggest that, if the concept of 'migraine' was familiar, a Romance term for it was not. Note the usage of the paraphrase *dolor de*/*en la media cabeça* to explain the concept to readers who may have been unfamiliar with the specialized terms used to identify it:[142]

[142] The generally accepted meaning of *cefalea* is usually broader (i.e., 'headache'), as demonstrated in *Gordonio* (1495): *dolor* (…) *en todo el cuerpo dela cabeça*: (…) *cephalea* 'pain throughout the entire head: cephalea.'

(1) *la dolor que se faze en la media cabeça a que llaman en arauigo* ***xaqueca****. and en latin* ***migranea***
'The pain that occurs in one half of the head that in Arabic they call xaqueca and in Latin migranea'
Lapidario, c. 1250

(2) *del dolor de la media cabeça; a que llaman en griego cefalea*
'of the pain in one half of the head that in Greek they call cefalea'
Lapidario, c. 1250

In contrast, the author identifies the broader meaning of 'headache' as *dolor de la cabeça* without explanation or an alternative signifier, presumably due to the transparency and familiarity of the term:

(3) [*la piedra*] *sana de la dolor de la cabeça; and del roydo que se faz en ella*
'the stone cures [one] of headache and of the noise that occurs in the head'
Lapidario, c. 1250

These examples suggest that Romance speakers and writers used a phrase, *dolor de cabeza*, to identify 'headache' but that the terms used for 'migraine' were not inherited, given the identification of the source languages of *xaqueca* and *migranea* alongside explanations of their meaning. As we explained in the previous chapter, if there is no native word for a particular concept (e.g., 'migraine'), then the introduction of a borrowing (e.g., *jaqueca*) is encouraged (Weinreich 1968).

5.2.3.1 Semantics of *jaqueca*

As the original source of Sp. *jaqueca* ‘migraine’ and Hispano-Ar. (*iš*)*šaqíqa* ‘migraine,’ Ar. *šaqīqah* ‘pain that affects one side of the head, migraine’) is abundant in Arabic language medical texts, with a relatively consistent core definition throughout its documentation:

(4) ***al-šaqīqa[h]*** *(jaqueca), afecta a la mitad de la cabeza, porque el cerebro está dividido en dos secciones. Por tanto, si el excedente tiende hacia el lado derecho, el dolor se produce en esta zona; mas, si aquel se inclina hacia el lado izquierdo, el dolor afectará a este lado.*
‘al-šaqīqa[h], affects half of the head, because the brain is divided into two sections. Therefore, if the pain is focused on the right side, the pain is produced in this zone; but, if it is focused on the left side, the pain will affect this side.’

Kitāb al-Dajīra fī ’ilm al- ṭibb, a 901
Ṭabit ibn Qurrá

(5) ***al-šaqīqa[h]*** *(jaqueca), tiene lugar dentro del cráneo, porque el cerebro se divide en dos partes* (…)*. Se trata igual que cualquier clase de dolor de cabeza.*
‘al-šaqīqa[h], occurs within the cranium, because the brain is divided into two parts. It is treated as any type of headache.’

Kitābu-l-Ḥāwī fī-1-ṭibb, a 925
Al-Razī

(6) ***Šaqīqa[h]*** *(jaqueca), es un dolor que abarca al oído, a la mitad de la cabeza y a parte del rostro*
‘Šaqīqa[h], is a pain felt in the ear, one half of the head, and part of the face’

Kitāb al-Manṣūrī, a 925
orig. Al-Razī

In these first examples from the tenth century, Ṭabit ibn Qurrá (a 901) and Al-Razī (a 925) highlight different aspects of migraine in *Kitāb al-Dajīra fī ’ilm al- ṭibb* and *Kitābu-l-Ḥāwī fī-1-ṭibb*, respectively, but they both share the broader definition of (*al-*)*šaqīqah*, that is, a type of headache that affects one half of the head. In *Kitāb al-Manṣūrī*, Al-Razī associates the additional pains of the ear and the face with migraine as well.

In the eleventh century, Albucasis (c. 1000) provides a definition similar to those of Ṭabit ibn Qurrá and Al-Razī, identifying the characteristic pain of migraine on one side of the head but also noting pain to the eye of the affected area:

(7) *Chapter three. On the cauterization of non-chronic migraine (**al-šaqīqa[h]**). When there occurs pain with headache in one side of the head and the pain extends to the eye*

Albucasis: On Surgery and Instruments, c. 1000
Albucasis

Avicena (c. 1020) offers a much more detailed description of migraine (including its cyclical nature, causes, other associated symptoms), but at its core the description of migraine is the same: a pain affecting one side of the head:

(8) ***al-šaqīqa[h]*** *(jaqueca): Decimos: es un dolor que afecta a uno de los dos lados de la cabeza. Galeno la define como el dolor que afecta a la zona intermedia. Tiene su origen o en el interior del cráneo o en la membrana que lo recubre o, muy frecuentemente, en el músculo temporal. Si es debida a causa externa, se es incapaz de soportar, el contacto. Además, las materias llegan hasta la zona, bien procedentes de las venas y arterias externas, bien del mismo cerebro y sus membranas... o puede ser causada, también, por vapores procedentes o de todo el cuerpo o de un órgano de esa parte. La **šaqīqa[h]** (jaqueca), suele ir acompañada de ciclos y, la mayor parte de las veces, tiene su origen en los humores. **Al-šaqīqa[h]** (jaqueca) no sobreviene por alteración de complexión simple, sino de humores, tanto cálidos como fríos, vientos o vapores. Sus síntomas son conocidos: en el frío, hallar alivio en el calentamiento, y distensión rápida. En el cálido, calidez al tacto, punzadas en las sienes y alivio con los refrigerantes. También percibir frío en el primero y calidez en el segundo, cuando el dolor aumenta.*
‘**al-šaqīqa[h]** (jaqueca): We say: it is a pain that affects one of the two sides of the head. Galen defines it as the pain that affects the intermediate zone. It has its origin either in the interior of the cranium or in the membrane that covers it or, very frequently, in the temporal muscle. If it is due to an external cause, one is unable to tolerate contact. Also, the matter arrives at the zone, either from external veins or arteries, or from the brain itself and its membranes, or it may be caused also by the wet humors from either the entire body or from a related organ. La **šaqīqa[h]** (jaqueca) tends to occur in cycles and, the majority of time, has its origin in the humors. **Al-šaqīqa[h]** (jaqueca) does not occur from a simple change in complexion, but rather due to the humors, both hot and cold, dry and wet. Its symptoms are known: in the cold, finding relief in keeping warm, and fast-onset strain. In the heat, warmth to touch, stabbing pains in the temples and relief with cooling. Also, feeling cold in the first case and warmth in the second, in which the pain increases.’

Al-Qānūn fī-l-ṭibb, c. 1020

Ibn Sīnā (Avicena)

Even in the late medieval period, as in the fourteenth-century texts of Ibn al-Jaṭīb (a 1374), the term is largely unchanged, focusing again on the localization of pain in one half of the head:

(9) *Es un dolor de cabeza circunscrito a la mitad de la cabeza*
'It is a headache limited to one half of the head'
Kitāb 'Amal man ṭabba li-man ḥabba, a 1374
Ibn al-Jaṭīb

(10) *Otro dolor de cabeza es el llamado* ***al-šaqīqa[h]****, que se localiza en la mitad de la cabeza y tiene su origen en los cuatro humores*
'Another type of headache is that called al-šaqīqa[h], which is located in one half of the head and has its origin in the four humors'
Urŷūza fī-l-ṭibb, a 1374
Ibn al-Jaṭīb

The first appearance of *jaqueca* in Castilian documentation is in the *Lapidario* (c. 1250 *xaqueca*), where its definition echoes those of the Arabic language texts:

(11) *la dolor que se faze en la media cabeça a que llaman en arábigo* ***xaqueca***
'the pain that occurs in one half of the head that in Arabic they call xaqueca'

However, there is debate regarding whether *jaqueca* may be considered an Arabism here or simply a translation. Neuvonen (1941) and Corominas and Pascual (1980-1991) present this first appearance as a translation; as such, it would not form a part of the Castilian lexicon at that time. However, even though Alfonso X indicates his awareness of the origin of the word, that fact

alone does not necessarily indicate that the word had not already been in use in Spanish at the time, as Herrera and Vázquez de Benito (1985) suggest.[143] In either case, the first documented orthographic variant of *jaqueca* is still *xaqueca*, which later resurfaces more than a century later in Alfonso Chirino:

(12) *Del dolor de cabeca and enla* ***xaqueca*** *que es dolor dela meytat dela cabeça*
'About headache and migraine, which is a pain of one half of the head'
Menor daño de la medicina, a 1429

This general definition of *jaqueca* (i.e., 'pain on one side of the head, migraine') is widespread through the end of the time period studied, as in the following example from Gómez Miedes:

(13) ***xaqueca****, que atormenta el medio casco de la cabeça por la distensión de las membranas*
'migraine, which torments one half of the head due to distention of the membranes'
Enchiridion o manual instrumento de salud contra el morbo articular que llaman gota, 1589

However, in one case from an anonymous author, *jaqueca* is described as a 'very serious headache.' While migraine may be considered a serious affliction, it is unclear from the context if the author is defining the term or simply describing it:

(14) ***jaqueca****, que es un dolor de cabeza muy grave*
'migraine, which is a very serious [type of] headache'
Deposición de la hermana Teresa de Jesús, 1596

[143] Herrera and Vázquez de Benito (1985) offer the contemporary example of "bocadillo en inglés se llama sandwich," in which the translated word is also one that has been fully incorporated into the borrowing language. Identification of the foreign origin of a word alone does not negate its incorporation into the language in question.

Given the rare instances in which authors equate *jaqueca* with the broader meaning of 'headache,' we will assume that the above example is an evaluation of migraine, rather than a definition. The distinction between *jaqueca* and the semantically broader *dolor de cabeza* and *cefalea* is maintained through the seventeenth century, as in Méndez Nieto, Vázquez de Espinosa, and Cobo, respectively:

(15) *Trata una breve y çierta cura de caephalea o* ***axaqueca****. Es la* ***axaqueca*** *enfermedad diuturna*
'Concerning a quick and certain cure for headache or migraine. Migraine is a lasting illness'
Discursos medicinales, 1606-1611

(16) *es muy conocida para dolores de caueça*[,] ***xaqueca****, y otras enfermedades*
'it is well known for headaches, migraine, and other illnesses'
Compendio y descripción de las Indias Occidentales, 1629

(17) *esta raíz* (...) *causa dolores de cabeza, vaguidos y* ***jaqueca***
'this root causes headaches, dizziness, and migraine'
Historia del Nuevo Mundo, 1653

5.2.3.2 Semantics of *hemicránea*

Like *jaqueca*, *hemicránea* first appears in the thirteenth century (c. 1250 *migranea*) but does not resurface until the fifteenth century in Spanish language medical texts. In many cases, patrimonial *hemicránea* is employed in much the same way as its synonym *jaqueca*, highlighting the unilateral component of the affliction, and often is used within the same sentences for the sake of clarification:

(18) ***Emigranea*** *es dolor en la meatad de la cabeça*

‘Emigranea is a pain in one half of the head’
Compendio de la humana salud (CHS), 1494[144]

(19) ***Emigranea*** *o* ***axaqueca*** *es delor* [*sic*] *dela meytad dela cabeça*
‘Emigranea or axaqueca is a pain in one half of the head’
Gordonio (GOR), 1495

(20) *Si es en meytad de la cabeça llamase* ***emigranea*** *que es* ***axaqueca****.*
‘If it is in one half of the head, it is called emigranea, which is axaqueca [migraine].’
Gordonio (GOR), 1495

(21) ***Emigranea*** *es en media cabeça vn dolor de dentro del craneo*
‘Emigranea is a pain within the cranium, in one half of the head’
Sumario de la medicina (SUM), 1498

(22) *demuestra vna enfermedat que dizen* ***evigrançe*** *.i.* ***enugeça*** *que es dicha enfermedat de la cabeça*
‘it demonstrates an illness that they call evigrançe and enugeça, which is an illness of the head’
Libro de recetas (RES), a 1500

(23) *E la dolor que faze en la media cabeça que laman* ***engranea***
‘And the pain that occurs in one half of the head that they call engranea’
Tratado de patología (TRM), a 1500

General unfamiliarity with the Latinism leads certain authors, such as Díaz de Isla in *Tractado contra el mal serpentino*, to continue the juxtaposition of *hemicránea* and *jaqueca* well

[144] Text titles followed by three-letter abbreviations are works found in the Spanish Medical Texts portion of the Digital Library of Old Spanish Texts. We have included the abbreviations for clarity as some of these medical texts have been assigned more than one name in different corpora.

into the sixteenth century. Elsewhere, the anonymous author of *Repertorio de los tiempos* maintains the distinction between *migranea* ‘migraine’ and *dolor de la cabeça* ‘headache’:

(24) *Si son dolores por gota* (...); *si es dolor de cabeça por* ***milgrania***[,] *por* ***axaqueca***
‘If they are pains, they are due to gout: if it is a headache, due to migraine’
Tractado contra el mal serpentino (YSL), 1542

(25) *En el medio de la cabeça está una vena; vale a* ***migranea*** *antigua y al dolor de la cabeça.*
‘In the middle of the head is a vein; it corresponds to old migranea [migraine] and to headache.’
Repertorio de los tiempos, 1554

5.2.3.3 *Migraña*

We also recognize the existence of a third term for ‘migraine’: *migraña*. Some sources (e.g., Corominas and Pascual 1980-1991: s.v. *cráneo*) claim that *migraña* is a semipopular term derived from Late Lat. HEMICRANIA. Under the heading of *axaqveca*,[145] Covarrubias (1611: s.v. *axaqveca*) instead identifies *migraña* as a word belonging to the Catalan of Valencia. This claim is reasonable given the similarity between Catalan *migranya* (first documented 1460) and Castilian *migraña*. Dworkin (2012) states that following the union of the Kingdom and Castile and Leon and the Crown of Aragon (1479), Castilian accepted loanwords in contact with the closely related languages of the latter territory (i.e., Aragonese and Catalan) that came to coexist in the new, unified state. While this is possible, the identification of a loanword as a Catalanism (as opposed to a Gallicism) in Castilian is often difficult to confirm.[146] If *migraña* is indeed a Romance borrowing, its introduction after the end of the thirteenth century more likely suggests a Catalan, rather than Gallic origin (Dworkin 2012).

[145] The variant *axaqveca* is not documented in CORDE.
[146] See Dworkin (2012) for further discussion of the difficulties in identifying loanwords as Catalanisms.

Although *migraña* is addressed in dictionaries in the early seventeenth century, it is absent from CORDE until the twentieth century,[147] suggesting that the term is not especially relevant for the present case study, but the term is an interesting case of lexical variation, nonetheless. The motivation for the introduction of *migraña* is unusual in that there had already existed two other equivalent terms for 'migraine,' one popular and the other learned. One possible explanation is the unfamiliarity of speakers with *hemicránea*, reflected in the scarce representation of the term throughout the documentation. During the time period in which *hemicránea* and *migraña* documentation overlap (i.e., twentieth century), *hemicránea* presents very few tokens in either CORDE (n = 5) or CREA (n = 4), where *migraña* is significantly more frequent in usage (n = 13, n = 181, respectively).[148] General speaker unfamiliarity with the original specialized term for 'migraine' may explain its replacement with another term (i.e., *migraña*) that would serve in the same capacity.[149] Curiously, *migraña* is now documented more than *jaqueca* and *hemicránea* combined and is employed as both a specialized (i.e., clinical) and quotidian term, suggesting that *migraña* may be taking over the functions of both *jaqueca* and *hemicránea*.[150]

5.2.3.4 Overview of semantics

[147] There are two examples of the form *migraña* in 1494, but we question whether this is the same *migraña* that is not again documented until the twentieth century. Given their chronology, these first few cases instead appear to represent one of the many popular forms that we have identified as belonging to medieval *hemicránea*, aware that the Latinate form is not in use until the nineteenth century or later.
[148] Our search was limited to documents from Spain.
[149] The *DLE* currently identifies *migraña* as a term belonging to the domain of medicine. In contrast, *jaqueca* is not assigned this designation.
[150] The evolution of the semantic category of migraine and similar lexical groups in modern medicine may be of interest for further investigation in the study of lexical variation.

From the documentation of *al-šaqīqah* in tenth-century Arabic texts through that of *jaqueca* in early modern Spanish documentation, authors have highlighted different aspects of migraine (to varying amounts of detail), but what nearly all examples have in common is its principal symptom, that is, pain (i.e., headache) that affects one side of the head. This core description of migraine is also found accompanying its Romance equivalent *hemicránea,* from its first incorporations into the written register to its modern-day use. With an overlap in both meaning (i.e., pain in one side of the head; migraine) and chronology (i.e., primarily in the fifteenth and sixteenth centuries but also later), *jaqueca* and *hemicránea* may be considered variants of the lexical variable 'migraine.'

Unlike the other two Arabisms that we have examined at length (i.e., *aceite* and *atalaya*), *jaqueca* has not undergone any significant semantic changes since its introduction into Spanish, nor have there been any observable semantic differences between the Arabism and its Latinate counterpart. One possible explanation for the close synonymy between *jaqueca* and *hemicránea* is that *jaqueca* (literally 'side or half of a twin object') is a calque of Gr. *hēmikranía* (Corriente 2008: 338), the original etymon of Sp. *hemicránea*. which may explain the very close synonymy of *jaqueca* and *hemicránea*. Given the tendency of authoritative dictionaries throughout the documentation to define *jaqueca* or *hemicránea* in terms of the other suggests that we cannot resort to differences in meaning to account for the preservation of multiple terms with the meaning of 'migraine.' After all, redundancy is generally unfavorable to the maintenance of Arabisms.

Other retained medical Arabisms present different outcomes. In general, maintained Arabisms that refer to medicine, remedies, and illnesses are those that have meanings that are somewhat different in denotative meaning from Latinate counterparts or that do not have counterparts at all. The divergent meanings between the Arabism and Latinate term are the result of semantic change, while other Arabisms identify a concept previously unknown to Hispano-Romance: *alfombrilla* 'measles' > 'rash, different from measles for the lack of catarrhal phenomena,' *hachís* 'Indian hemp' > 'hashish,' *zaratán* 'cancer' > 'breast cancer,' as well as *momia* 'mummy,' *zaragatona* 'fleawort (plant)'. For this reason, then, in terms of semantics, then, the retention of *jaqueca* is not typical of the treatment of medical Arabisms of the time period. Although the absence of semantic differences between *jaqueca* and *hemicránea* does not favor retention of the Arabism, we will now turn our attention to sociolinguistic factors, which will prove to be more successful in this regard.

5.3 Sociolinguistic factors

Having explained the extent to which our selected linguistic factors have favored the maintenance of *jaqueca* in section 5.2 (formal variation, very favorable; related vocabulary, favorable; semantic differences, unfavorable), we may now determine the significance of the sociolinguistic factors of dialect (section 5.3.1), register (section 5.3.2), and language contact (section 5.3.3). In many cases regionalized usage, differences in usage across register, and neutral or positive attitudes toward the semantic field in question favor the retention of an Arabism. In the case of *jaqueca*, as we will demonstrate below, two of these factors are favorable while the other is irrelevant.

5.3.1 Dialect

Having addressed linguistic factors, we may now focus on the role of sociolinguistic factors in the maintenance of *jaqueca*. Once more, the reader will recall that dialectal differences tend to be significant in sociolinguistic research, especially when Arabisms are involved. The existence of regional Arabic-based vocabulary is the result of the greater duration of contact between Arabic and Romance in the southern regions of the Iberian Peninsula, where such vocabulary has the greatest vitality and is highest in number. In many cases, this has led to a certain degree of diglossia for particular lexical items, where speakers of a given region (usually those in the South) employ a Romance or Latin-based word as the standard term alongside an Arabic-based word that exists regionally, but that has little currency outside the region in question.

As in the previous examples of *aceite* and *atalaya, jaqueca* is a term that is used in both European and Latin American Spanish. In this regard, it is similar to the relatively small number of Castilian Arabisms in the domain of medical terminology that have survived past the Middle Ages. For example, in a series of articles concerning medieval medical terms of Arabic origin, Herrera and Vázquez de Benito (1981, 1982, 1983, 1985) have addressed nearly 50 such lexical items, with only seven appearing in the most recent edition of the *DLE*, and only four of these are considered contemporary.

Neuvonen (1941: 310) explains that the absence of learned Arabisms in the first stage of borrowing of Arabisms into Hispano-Romance (eighth through eleventh centuries) may have been due to their fleeting nature. Perhaps a more precise explanation of this phenomenon is that

provided by García González (2007), that most learned Arabisms from this first period are rarely documented and many are found only in a single document (i.e., a single author). If these Arabisms were only used by certain individuals (e.g., Andalusi Romance speakers who fled al-Andalus), their scarce documentation may provide evidence that these words had not established themselves in the best-known Castilian Romance varieties of the North, which explains their failure to establish themselves in Castilian.

In the second stage of borrowings of Arabisms into Spanish (twelfth and thirteenth centuries), García González (2007) identifies a considerable number (n = 32) of scientific Arabisms belonging to the disciplines of astronomy, botany, medicine, and others. However, the figure is not as significant as one might expect since a large portion of the words are either poorly integrated, lost, or accidental Arabisms (i.e., appearing only once). Many of these terms entered into Castilian through the written register; again, we have to question the degree of integration, if any, of these Arabisms into Castilian at the time. That is not to say that that applies to all scientific terms documented during this period: *Jaqueca*, for example, is one Arabism that, despite a singular documentation in the thirteenth century, becomes widespread in use by the fifteenth century.

As we stated above, there are only seven medical Arabisms discussed by Herrera and Vázquez de Benito (1981, 1982, 1983, 1985) that appear in the *DLE*: these are *aliacán* 'jaundice,' *botor* 'bubo,' *cifaque* 'peritoneum,' *hadruba* 'hump,' *jaqueca*, *safena* 'saphenous (vein),' *zaratán* 'breast cancer.' Of these, only four (i.e., *aliacán*, *jaqueca*, *safena*, *zaratán*) are considered contemporary. Since *hadruba* and *botor* appear in few texts in the documentation, a

probable explanation is that these are Arabisms that were never well established in the Castilian speech community. On the other hand, medical Arabisms that endured into the modern age are recorded in numerous texts, although to different degrees. What they have in common is that none of the existing medical terms addressed by Herrera and Vázquez de Benito (1981, 1982, 1983, 1985) (i.e., *aliacán*, *safena*, *zaratán*) are regional. However, given their sparse appearances in contemporary texts, *aliacán* and *zaratán* appear to have waning vitality, no doubt the result of the preference for Latin, Greek, and hybrid Latin-Greek terminology in the domain of medicine. As the only term available for the meaning of 'saphenous (vein),' *safena*, which also has its ultimate origin in Greek, has a secure place in the medical lexicon (Gutiérrez Rodilla 2014). In this way, the Arabic-based *jaqueca*, which shares its meaning with other words of Latin-Greek origin, is more similar to *aliacán* and *zaratán* than *safena*. Whether *jaqueca* will eventually be eliminated by *hemicránea* or *migraña* is uncertain, but what we can confirm is that dialectal differences are not a relevant factor in the maintenance of this Arabism.

5.3.2 Register

Once more, we will determine the extent to which register may have played a role in the preservation of Arabisms. In section 5.2.3, we claimed that *jaqueca* and *hemicránea* were very close in meaning, with neither term deviating from the core meaning of 'migraine.' Of course, certain authors provide additional (i.e., secondary) symptoms, but in both cases the central meaning is the same, that is, 'pain in one side of the head,' the characteristic symptom of migraine. In this sense, the individual histories of these words in the Spanish documentation are relatively static, with no significant observable semantic changes occurring in either term. The relevant question, then, is whether these terms differ in other ways, for example, in register. As

in the previous chapters, we will establish if register is a significant factor in the retention of the Arabism through an examination of the text types in which *jaqueca* and *hemicránea* are found in the documentation of the time period in question.

5.3.2.1 Initial documentation of *jaqueca* and *hemicránea*

Both *jaqueca* and *hemicránea* are first documented in CORDE in the mid-thirteenth century. Due to the paucity of tokens of either term at this time, *jaqueca* and *hemicránea* are found only in the text category of scientific prose. It is not until two centuries later that both terms will firmly establish themselves in the documentation.

5.3.2.2 Coexistence of *jaqueca* and *hemicránea*

In the fifteenth century, after some two centuries of absence, *jaqueca* and *hemicránea* reappear in the documentation as variants of the same lexical variable, 'migraine.' At this time, *jaqueca* is found in three text types overall, namely scientific prose (68.42%), narrative prose (21.05%), and lyrical verse (10.53%). Although *hemicránea* also favors scientific prose (100.00%), it is less varied with regard to text category, being found only in one type. Due to the nature of the text types in which both terms are found, as well as the texts themselves (i.e., medical texts, in large part), *jaqueca* and *hemicránea* are found entirely in high register texts at this time. There are no tokens of either term that appear in texts that reflect speech or popular language, a result that contrasts with those of the other Arabisms that we have examined in this study. In other words, there is no evidence of a distinction in register between *jaqueca* and *hemicránea* in the fifteenth century.

In the following century, usage of *jaqueca* expands into other categories, as made evident by its presence in scientific prose, lyrical verse, narrative prose, historical prose, dramatic prose, religious prose, didactic prose, dramatic verse, legal prose, and societal prose. Although scientific prose (27.27%) is still the preferred text type for *jaqueca* in the sixteenth century, authors use *jaqueca* in an array of text types. What is even more notable for the Arabism at this time is that it appears in a considerable number of texts (21.21%) that reflect popular language. On the other hand, *hemicránea* is still found only in high register texts of scientific prose, although its frequency (n = 2) is considerably lower than that of *jaqueca* (n = 33).

Although rare in the sixteenth century, *hemicránea* is completely absent from the documentation in the seventeenth century. In this time period, *jaqueca* flourishes as the only term we encountered in the Spanish language documentation for 'migraine,' appearing in many of the same text categories as it had in the previous century. However, unlike the previous century, scientific prose (19.23%) is no longer the most frequently recorded category for *jaqueca*, but it is instead narrative prose (25.00%), followed by lyrical verse (23.08%). We also found that *jaqueca* was documented less in popular register contexts (9.62%) than in the previous century, but the absence of *hemicránea* from popular contexts until this point in time is suggestive of the continued preference for *jaqueca* in the popular register.

Through the seventeenth century, at least, *jaqueca* is represented in a number of different text categories, whereas *hemicránea* is restricted to the category of scientific prose. Not only does *jaqueca* have representation in numerous text types, but it is also the only term of the two that has demonstrated examples from the popular register. Due to the difference in representation

of *jaqueca* and *hemicránea* across register (i.e., the appearance of *jaqueca* in texts that represent popular language or speech and the complete absence of *hemicránea* in such texts), *jaqueca* is clearly the preferred term for 'migraine' in popular usage (and perhaps the only one until the introduction of *migraña*). The fact that *jaqueca* is the only term available for 'migraine' in popular usage, and that *hemicránea* is used exclusively in high register texts, likely favored the retention of an Arabism (i.e., *jaqueca*) that had to compete with a synonymous patrimonial term (i.e., *hemicránea*).

After examining all the medical Arabisms addressed by Herrera and Vázquez de Benito (1981, 1982, 1983, 1985) and Maíllo Salgado (1998), there are very few cases where a retained Arabism coexists with a learned term that is truly synonymous with it. As we revealed above in section 5.2.3.3, most maintained medical Arabisms differ slightly in meaning to closely related learned terms. Other medical Arabisms, those that faced competition with a true synonym, were usually lost, as in the case of *botor* (vs *buba*) and *cifaque* (vs *peritoneo*). What these lexical pairs have in common is that all four terms are found almost exclusively in medical texts, suggesting that all terms were generally restricted to high register usage, a factor that is unfavorable to the retention of the Arabism. On the other hand, one retained Arabism that we found to have coexisted with a learned counterpart might have been distinguished by register: *aliacán* (vs *ictericia*), although documentation of *aliacán* is too rare to make a definitive statement in this regard.[151] With these facts in mind, the documentation supports that medical Arabisms that are not unique in meaning are generally lost, unless they are (or become) distinguished by usage across register. In this regard, *jaqueca* once again is treated in a different manner from other

[151] We must also question the continued vitality of *aliacán*, as it is rare in CORDE, absent from CREA, and is used only metalinguistically in a single case in CORPES XXI. It may be relegated to informal speech.

medical Arabisms, with *jaqueca* established as the quotidian term and the Latinate equivalent reserved for learned use.

5.3.3 Language contact

The reader may recall from the second chapter (section 2.3.4.3.3) that the context in which contact between Arabic and the different stages of Romance took place contains information that is essential for our understanding of the maintenance of words of Arabic origin in Castilian. Specifically, we claimed that the means of transmission may have influenced the survival of Arabisms. Again, by categorizing such words as either direct or indirect, we hypothesized that indirect Arabisms would be more resistant to loss than direct ones. Another aspect that we have examined is the awareness on the part of Spanish speakers regarding the Arabic origin of the word in question. Each of these components of language contact will be treated individually in the sections below.

5.3.3.1 Transmission of medical Arabisms in general and *jaqueca* in particular

Although the contribution of Arabisms to Ibero-Romance languages that occurred through contact between Arabic and Romance in the informal, spoken register is significant, another important medium of transmission for Arabisms, particularly medical terms, is the written register. In numerous cases, Arabic medical terms entered into Romance languages through translations. Of course, the practice of translation relies on the existence of equivalent terms or structures in the target language; however, it is the absence of such equivalents that is perhaps more conducive to the borrowing of foreign elements, which is precisely the case of the

borrowing of medical Arabisms originating in the written register into Spanish and other Romance languages.

Before the thirteenth century, Castilian Romance had been primarily a spoken language, but once scholars began to translate languages (e.g., Arabic) with established written registers Castilian Romance had to adapt, adopting morphological and syntactic structures and, of course, vocabulary (Winet 2006: 200). This effort of expansion in these areas (i.e., morphology, syntax, vocabulary) would allow writers to express in Spanish complex ideas in the various domains of medieval knowledge (Bossong 1982: 1-11). In the case of vocabulary, voids in the lexicon were often remedied by individual borrowings.

If the borrowing of Arabisms through speech had its complications (i.e., diverse phonological representations), then the transmission of Arabisms through translations was problematic in its own right. In addition to the difficulty of Romance speakers in interpreting the Arabic article *al-* as part of the term or not (e.g., *alxaqueca* vs *xaqueca*), which it shares with terms transmitted through speech, Arabisms incorporated from translations have other interpretive challenges. One of the more obvious is that of assimilation. In speech, sun and moon letters in the root govern whether the /l/ of the article is phonetically assimilated or unassimilated, respectively. Transcription, on the other hand, led scribes to produce assimilated and unassimilated versions of a given word with a sun letter (e.g., *axaqueca* vs *alxaqueca*), which in speech would have demanded assimilation of the consonant (Winet 2006: 199). Finally, errors in interpretation of the letters themselves led to variable outcomes, as in the case of OSp.

cinzen, sirçen, sirsen, sirzen, xerçi ‘brain tumor; frenzy’ or *alchoboin, althohoin, althoin, althaun* ‘plague, epidemic.’

Such difficulties in the transcription of Arabic medical texts has led many medical Arabisms to acquire multiple orthographic variants throughout its documentation. However, the variants themselves are often highly informative. The earliest variants of *jaqueca* (i.e., those from the medieval period) are consistent in that they all contain *x*, as in *alxaqueca, axaqueca,* and *xaqueca*. Herrera and Vázquez de Benito (1985) indicate that the transliteration of the letter *šīn* (i.e., ش, representative of /ʃ/) as *x* in these three variants in Old Spanish suggests an initial diffusion of *jaqueca* through speech, in contrast to many other (learned) medical terms of Arabic origin that first passed from Arabic to Latin before entering into Hispano-Romance. Due to the lack of the phoneme /ʃ/ in Latin, Arabic terms that were transliterated into Latin would have been represented as *s*. On the other hand, /ʃ/ was equivalent to Old Spanish *x,* which explains the use of *x* in the Castilian equivalent, *xaqueca*. Following the simplification of the phonological system from Old Spanish to Modern Spanish, /ʃ/ is replaced by /x/, which explains the appearance of modern *jaqueca*.

Based on its documentation (i.e., representation only in high register Spanish texts until the sixteenth century) and its domain (i.e., medicine), *jaqueca* is also a term with an established history in the written register. However, it is through speech that the Arabism is first transmitted, far surpassing the utility (and frequency) of its Romance counterpart *hemicránea*. In other words, *jaqueca*, which also penetrated the language in educated, written language, first established itself as an indispensable term in speech.

In other cases, Arabisms documented in medical translations are generally less successful. Consider, for example, the many terms documented in medical texts that are no longer in use: *anzarote, cifaque, jectigacion*, *mirach, silac*. These terms and many others were likely reserved, in large part, for the written register, where they competed with Latinate synonyms. Other terms found in medical texts, which were transmitted through speech (e.g., *alcatenes*, *alfombrilla*, *almorí*, *jarabe* / *jarope*, *zaragatona*, and of course *jaqueca*) are generally better maintained than their bookish counterparts.

5.3.3.2 Speaker awareness and attitudes

As with the other two Arabisms that we have studied, *jaqueca* and its variants follow the structure of the Spanish syllable, so that its phonological features are not especially revealing of its Semitic origin. Consider how the Arabism resembles other Spanish terms, both in its initial and final features. For instance, numerous popular and learned terms of different origins in Hispano-Romance end in the sequence -*eca*: nouns *apoteca* (< Lat. APOTHECA), *biblioteca* (< Lat. BIBLIOTHECA), *hipoteca* (< Lat. HYPOTHECA), *manteca* (of unknown origin), *muñeca* (of pre-Roman origin)[152] and the feminine forms of certain adjectives, such as *hueca* (der. of Lat. OCCARE), *intrínseca* (< Lat. INTRINSECA), *seca* (< Lat. SICCA). In other words, the presence of the ending -*eca* in Hispano-Romance is not only due to patrimonial terms (e.g., *seca*) and Latinisms (e.g., *apoteca*) but also from much earlier sources (e.g., *manteca*, *muñeca*). There are also non-Arabic-based words beginning in /(a)xak-/,[153] but they are uncommon or obsolete (e.g.,

[152] CORDE identifies the origin of *manteca* as uncertain, whereas that of *muñeca* as pre-Roman. Dworkin claims *manteca* is pre-Roman (2012: 31).
[153] The Old Spanish equivalent of this sequence is /(a)ʃak-/.

jaqués, *jaquir*). Instead, most of these are Arabisms: *ajaquefa* (< Hispano-Ar. *assaqífa*), as well as *jaque* (< Cl. Ar. *šāh*), *jáquima* (< Hispano-Ar. *šakíma*) and their numerous derivatives. However, the ability of the typical language user to identify an Arabism as such based on its beginning with /(a)xak-/ is questionable considering its proximity to patrimonial terms and intra-Romance borrowings that contain similar initial sequences: *aja* (< Lat. ASCIA), *ajar* (< *ahajar*, der. of Vulg. Lat. *FALLIA?), *ajo* (< Lat. ALIUM), *jacer* (< Lat. IACERE), *jaca* (< OSp. *haca* < OFr. *haque*), *jacobino* (< Fr. *jacobin*).

Beyond the level of phonology, speakers may have been able to identify *jaqueca* due to certain morphosyntactic characteristics. With regard to the morphosyntactic features of its variants, *jaqueca* is more similar to *atalaya* than *aceite* since both *jaqueca* and *atalaya* had alternating forms with the Arabic article (i.e., *a*(*l*)-) and without it. What distinguishes *jaqueca* from *atalaya*, though, is that both types of *jaqueca* forms coexist through the end of the period of interest; in the case of *atalaya*, only the form with the article survives beyond the fifteenth century. Still, forms without the article (e.g., *xaqueca*) are dominant throughout the time period studied, and even those with the article (e.g., *axaqueca*) are almost always assimilated, with examples with an unassimilated article (e.g., *alxaqueca*) representing only a fraction of the documented tokens and appearing only in a single document. Even if the forms containing the assimilated article were ultimately preferred, Arabisms with assimilated articles are not as obviously Arabic in origin as those with unassimilated articles. Once more, the initial *a*- of *ajaqueca* (let alone its absence in *jaqueca*) is not especially revealing of the origin of this Arabism.

Given that neither phonological nor morphosyntactic characteristics are particularly revealing of the Arabic origin of *jaqueca*, speakers might have been aware of the Arabic origin of the term on the basis of semantics. As in the case of military knowledge, there is also considerable transmission of medical knowledge via Arabic, (although the transmission of military and medical knowledge generally occurs through different media, that is, speech and writing, respectively). In the Middle Ages (especially the Late Middle Ages), transmission of knowledge relied heavily on the translation of texts into the Romance languages. Although Arabic (and other Romance language) texts containing specialized knowledge were translated into Castilian, the majority of translations were based on Latin texts. It is also at this time that writers began to compose Romance medical texts that were inspired, again, by mostly Latin (but also Arabic) sources (Sánchez González de Herrero and Vázquez de Benito 2010: 79). In addition to the smaller number of Arabic-sourced texts compared to Latin-based ones, there is the broader issue of accessibility of all medical texts, which, with their elevated, specialized language and vocabulary, may only have been accessible to certain restricted elite (Sánchez González de Herrero and Vázquez de Benito 2010: 82). With such restrictions, the broader speech community would have been unfamiliar with much of the newly introduced vocabulary as made evident by the fact that few medical Arabisms have survived the medieval period. On the other hand, *jaqueca* is a peculiar case since, unlike other Arabic medical terms, it has had considerable vitality in popular or spoken language, outside of the written register. Still, the greater contribution of Arabic to medical knowledge had to have been largely unnoticed. Therefore, to the greater speech community, semantic considerations are not revealing of the origin of *jaqueca* either.

Since neither the appearance nor meaning of *jaqueca* are particularly informative, it is likely that Spanish speakers were generally unaware of the Arabic nature of this Arabism. On the other hand, we must still consider the role of attitudes toward *jaqueca*. Given the general opposition of Spanish writers toward the use of Arabisms, we should assume that subjective references to *jaqueca* will be unfavorable.

If metalinguistic references to *atalaya* and its origin are scarce in the documentation, then those concerning *jaqueca* are virtually absent. In the time period studied, there is only one such example, the very first token of *jaqueca* in the documentation:

(26) *la dolor que se faze en la media cabeça a que llaman en arauigo* ***xaqueca****. and en latin migranea*
'the pain that occurs in one half of the head that in Arabic they call xaqueca and in Latin migranea'

Lapidario, c. 1250

By identifying the Arabic origin of *jaqueca* and its synonymy with *migranea*, the above example appears to be neutral, containing no obvious subjective remark regarding prescriptive usage or recommendation for a Romance equivalent.

While there are a few other pertinent examples of *jaqueca* throughout the documentation, they are not strictly metalinguistic uses. In its second documentation, *jaqueca* is accompanied by a gloss, which is used to explain a term that may have been unfamiliar to certain readers. The use of *jaqueca* in this example is didactic and clearly neutral in tone:

(27) *Del dolor de cabeca and enla* ***xaqueca*** *que es dolor dela meytat dela cabeça*
'About headache and migraine, which is a pain in one half of the head'
Menor daño de la medicina, a 1429

Other examples from the domain of medicine are indicative of neutral attitudes toward the use of *jaqueca* in the elevated register, even in scientific texts:

(28) *Si es en meytad dela cabeça llamase emigranea que es* ***axaqueca***
'If it occurs in one half of the head, it is called emigranea, which is axaqueca [migraine]'
Gordonio (GOR), 1495

(29) *no sólo el morbo articular, que vulgarmente llaman gota* (...) *al que llamamos* ***xaqueca****, que atormenta el medio casco de la cabeça*
'Not only articular disease, that they commonly call gota [gout], [but also] that which we call xaqueca, which torments one half of the head'
Enchiridion o manual instrumento de salud contra el morbo articular que llaman gota, 1589

The frequent usage of *jaqueca* (rather than *hemicránea*) in medical texts is surprising, given the demand of the emerging language of the scientific disciplines for precision (Gutiérrez Rodilla 2005: 22-23), an objective that was made difficult since many Arabisms had unknown etymologies (Giménez-Eguíbar 2011: 237). Despite its usage in popular language and Arabic origin, any commentary on *jaqueca* is objective or neutral, as it is generally employed alongside glosses (i.e., patrimonial equivalents). The use of *jaqueca* as the primary term for 'migraine' is well established, and if speakers (centuries after its identification as an Arabism by Alfonso X) were aware of its origin, there is no evidence in the documentation that they were opposed to the term.

If the documentation is a faithful reflection of attitudes toward the Arabism, then *jaqueca* is viewed neutrally. Such attitudes toward *jaqueca* may be attributed to another factor, that is, speaker awareness, since speakers must have been largely unaware of its Arabic origin, which is not immediately obvious on the basis of phonology, morphosyntax, or semantics. Viewed as an integral part of the lexicon, then, neutral speaker attitudes toward *jaqueca* are favorable to its maintenance in the language.

In contrast, a large portion of the other medical Arabisms documented in the medieval and early modern eras were more easily identified, and eliminated, as a result of their nonnative structure. Some of the many lexical items that are much more conspicuous in their foreign nature than *jaqueca* are *mirach* (especially variant *almirach*), *silac* (especially variant *alselach*), and *flisei* 'love' (especially variants *alhasch*, *ylischi*). Due to the development of learned vocabulary from Latin, Greek, and hybrid Latin-Greek in the scientific disciplines during the Late Middle Ages and early modern eras, many of these medical Arabisms must have been viewed unfavorably. As in the previous case studies, neutral attitudes toward the Arabism are a decisive factor for its retention.

5.4 Summary

In the present chapter, we have analyzed the effects that linguistic and sociolinguistic factors have had on *jaqueca*, as well as the context in which the term and others were transmitted. Of the six factors that we identified in this chapter, four are favorable to the retention of the Arabism. Of these, formal variation, register, and language contact appear to be most favorable, although related vocabulary is favorable to retention as well. Semantic

differences between *jaqueca* and *hemicránea* are not significant and, therefore, are an unfavorable factor. Finally, the factor of dialect, as in the other two case studies, is not significant in the maintenance of the Arabism.

In terms of formal variation, *jaqueca* is considerably less varied than its Romance synonym, *hemicránea*. In the documentation, there are two advantages that the Arabism has over its counterpart: on one hand, *jaqueca* orthographic variants are far fewer than those of *hemicránea*; on the other hand, the *jaqueca* variants recorded are far more similar in form to one another than those of *hemicránea*. As a term that flourished in the speech community, *jaqueca* was much more familiar to speakers than *hemicránea*, which may have been a term known only among those well versed in the medical sciences (Herrera and Vázquez de Benito 1985: 99). It is for this reason that the factor of formal variation here is informative beyond the factor itself; in other words, it suggests that *hemicránea*, although a specialized term in its own right (due to its usage as a medical term), has never been a significant competitor of the much more frequent and popular *jaqueca*, whose established position in the language is also due to another highly favorable factor.

According to the documentation available through the seventeenth century, representation of *jaqueca* and *hemicránea* across different text types is suggestive of the different uses that each of the terms developed over the centuries. The Arabism is easily the more varied of the two terms, appearing in various text categories, whereas the learned term is encountered only in scientific prose. Furthermore, *jaqueca* is employed in texts that denote popular language or speech, while *hemicránea* is entirely absent from such texts. In other words, there is a clear

division in register usage between the two terms through the time period studied, which in part explains the maintenance of the Arabism alongside a Latinate term. This argument is reasonable given that other such lexical pairs (whereby a quotidian term for a medical condition coexists with an equivalent, specialized medical term) are abundant across European languages (e.g., Eng. *headache* / *cephalalgia*; Sp. *fiebre* / *pirexia*).

Although semantic differences between *jaqueca* and *hemicránea* are not substantial, we must note another, significant semantic consideration, that of perceived necessity, one of the most common motivations for borrowing. Since there is no documentation of a patrimonial term for 'migraine,' Hispano-Romance speakers adopted a word (i.e., an Arabism) that would occupy that space. As a word introduced into Castilian Romance for that very purpose, *jaqueca* squarely falls into the category of cultural borrowing. In use in spoken language during the thirteenth century (and likely much earlier) *jaqueca* was long unchallenged by an equivalent quotidian term: it was only in the seventeenth century that *migraña* (< Cat. *migranya*) appeared in the documentation; however, by that time *jaqueca* had already established itself as the usual term for 'migraine.'

As we stated earlier in the study, many medical words of Arabic origin were lost from Hispano-Romance and, later, Castilian for having originated from a language and culture in decline. Adding to this the fact that *hemicránea* occupied the same semantic space as *jaqueca*, it would be reasonable to predict the demise of the Arabism alongside patrimonial *hemicránea*. However, various other factors, both linguistic and sociolinguistic, have proven to be favorable in its retention. Again, the history of *jaqueca* (and other medical Arabisms) is complex due to the

multiple factors involved in its retention, but through an analysis of multiple factors we may better comprehend its intricacies.

As an example of terminology in the field of medicine, remedies, and illnesses, *jaqueca* is somewhat unusual. Formal variation of *jaqueca* is moderate: Although it displays greater variation than other retained medical Arabisms (e.g., *zaratán* / *çaratán*, *momia* / *mumia*, or *quina*), its variants are relatively consistent in form in comparison with those of a large number of medical Arabisms that were eliminated, such as *alhebria*, *botor*, *mirach*, *silac*, *subet*, and *flesei*. Most medical Arabisms, whether maintained or lost, typically give rise to no more than one (and usually no) derived words, whereas *jaqueca* is prolific. The only other comparable medical Arabism is *jarabe* / *jarope* (*jarabear*, *jaropar* / *jaropear*, *jaropeo*, *jarapotear*); that is, only two of the more than sixty medical Arabisms that we examined contained more than one derived lexical item.[154] Another unusual feature of *jaqueca* among the majority of retained medical Arabisms is that it coexists with a synonym: most of these terms either (1) do not exist alongside a true synonym (e.g., *alcatenes* 'linen ointment,' *julepe* 'sedative drink,' *quina* 'cinchona (plant),' *zaragatona* 'fleawort') or (2) underwent semantic change to distinguish themselves from their counterparts (e.g., *alfombrilla* 'measles' > 'rash, different from measles in lack of catarrhal phenomena,' *zaratán* 'cancer' > 'breast cancer').

However, *jaqueca* is similar to other medical Arabisms in that it too is not restricted in its geographical distribution, that is, dialectal differences are generally not a relevant factor.

[154] Our analysis includes the medical Arabisms discussed by Herrera and Vázquez de Benito (1981, 1982, 1983, 1985) and Maíllo Salgado (1998). *Momia* has led to *momificación* and *momificar*, but these entries are documented well outside of the time period of interest (i.e., nineteenth century).

Another way in which *jaqueca* is treated differently than most other retained medical Arabisms is that there are few that coexist with true learned synonyms, with the Arabism typically eliminated in competition with its counterpart. Attitudes toward *jaqueca* are neutral, perhaps due to its conventional form (unlike many Arabisms in medicine), which may have concealed its origin. Finally, based on our analysis of the term and others in the same domain, it appears that transmission through speech (e.g., *jaqueca*, *jarabe* / *jarope*, *zaragatona*) may have been more favorable to the retention of medical Arabisms than transmission through writing (e.g., *mirach*, *sifac*, *silac*).

6 Conclusion

6.1 Introduction

The purpose of the present study was to further investigate lexical rivalries between Arabisms and Latinate or Romance terms, a linguistic phenomenon that has received little attention in the literature (cf. Eberenz 2006, Dworkin 2012). Even though there has been some recent interest in researching the loss of Arabisms in the face of competition from Latinate or Romance terms (cf. Giménez-Eguíbar 2011), the question of maintenance of Arabisms in similar competition has gone unnoticed. To address the absence of research in this area, we posed the following research questions: (1) What specific factors allowed Arabisms to coexist with Romance terms with which they are synonymous?, (2) What processes supported retention of old Arabisms and introduction of new Arabisms in the language at the same time that other Arabisms were ousted from Castilian?, and (3) More generally, why were certain Arabisms maintained or introduced when others were not?

Using a sociolinguistic perspective to investigate these questions, we directed our attention to both linguistic and sociolinguistic factors, examining how they could explain the retention of Arabisms, which generally were viewed negatively in the late medieval and early modern eras, leading to their obsolescence and eventual loss. Further complicating the matter of maintenance—and what makes it more intriguing—is the fact that many retained Arabisms persisted in spite of competition with other Latinate or Romance synonyms. Dworkin (2012) has briefly treated the phenomenon; however, in his study, lexical rivalries are treated as individual cases, meaning that the author does not provide broader conclusions for the retention of

Arabisms in competition with Latinate or Romance terms in any given domain or as a whole. The present study is a first step in that direction.

Below, we first revisit our observations from our introduction, our chapter of methodological and theoretical issues, and our three cases studies. These are followed by an examination of the generalizations that can be drawn regarding the treatment of Arabisms in a general sense, and finally we discuss considerations for future research in the realm of retention of Arabisms.

6.2 Historical background

In chapter one, we outlined the historical background of contact between Arabic and Castilian Romance in the medieval and early modern eras that encouraged borrowing from Arabic to Castilian Romance. Of these various types of borrowing (e.g., phonological, morphological, semantic, syntactic, lexical), lexical borrowing of Arabic is easily the most conspicuous in Castilian. The semantic domains covered by these lexical borrowings are diverse and include agriculture, the military, health and medicine, and many others. Although the presence of Arabic loans in Modern Castilian is still considerable, many loanwords introduced during the medieval period were eliminated due to negative attitudes propagated by the learned elite during the late medieval and early modern eras.[155]

[155] We discuss attitudes toward the Arabic influence of the Spanish lexicon in sections 1.1 (generally), 2.3.4.3.3 (as a sociolinguistic factor), 3.3.5.2 (agricultural Arabisms), 4.3.3.2 (military Arabisms), 5.3.3.2 (medical Arabisms). Giménez-Eguíbar (2011) treats the matter in greater detail, especially in sections 3.0-3.3.

Since research on Arabisms (i.e., Arabic loanwords) in Castilian has tended to focus on their elimination, it has neglected to explain why these words are maintained or incorporated. To address this void in the literature, we analyzed the historical development of three representative sets of competing Arabic-Romance synonyms using methodology from traditional historical linguistics and sociolinguistics, an approach developed by scholars of the relatively new linguistic subdiscipline of historical sociolinguistics. The semantic categories selected for analysis are diverse and include terminology in the domains of oleiculture (*aceite* / *óleo* / *olio*), the military (*atalaya* / *centinela*), and health (*jaqueca* / *hemicránea* / *migraña*).

6.3 Our approach to the study of Arabisms

In chapter two, we explained how other researchers approached language variation in general, as well as how the concept of the linguistic variable has evolved—after some debate—so that it now may be used to analyze variation occurring in levels of language beyond phonology and morphology, including syntax and the lexicon.[156] We agree with Labov (1978) that the linguistic variable can be a useful tool to study lexical variation, and that historical variation can be studied from a sociolinguistic perspective. Since lexical variation is a more recent subcategory of language variation, we found it necessary to describe methodological and theoretical issues that may arise in studies of lexical variation in general and of the study of Arabisms in particular. Once we addressed these difficulties, we reviewed how sociolinguistic researchers have addressed the role of different linguistic (internal) and sociolinguistic (external) factors in lexical variation. Based on the findings of previous studies, lexical variation may be

[156] Section 2.2 addresses the debate between Lavandera (1978) and Labov (1978) over the utility of the linguistic variable for research in variation beyond the level of morphology. The reader will recall that former author questioned its utility, while the latter author explained why it is a legitimate tool for studying other levels of language, including the lexicon.

favorable or unfavorable to the retention of words that are synonymous with others, depending on the type of variation in question (i.e., the different factors).

The factors that we selected are the following: the linguistic factors of formal variation, related vocabulary, semantic differences, as well as the sociolinguistic factors of dialect, register, gender, age, and language contact. In our study, we determined that limited formal variation, a well-established group of related (and/or derived) vocabulary, and semantic differences between the Arabism and the Latinate or Romance term(s) were all favorable to the retention of the Arabism. We also predicted that regionalized usage, differences in usage across register, gender, and age, and neutral or favorable attitudes would favor the Arabism.

6.4 *aceite* and agricultural Arabisms

In chapter three, we addressed the behavior of the lexical group of *aceite, olio*, and *óleo*, as well as the treatment of other agricultural Arabisms in Castilian. Of the linguistic and sociolinguistic factors that we addressed, the most favorable to the maintenance of *aceite* are semantic differences and the circumstances of language contact, specifically neutral attitudes and intensity (that is, duration) of contact.

The introduction of the Arabism was encouraged by a perceived lexical void, or vacancy, to refer to a new material reality (i.e., olive oil) (Lodares 1992, Cano Aguilar 1993). Through experimentation, speakers began to use *aceite* with increasingly broader uses, to the degree that it covered much of the same semantic territory that *olio* had initially. Its patrimonial counterpart, on the other hand, was gradually used by speakers in increasingly restricted contexts, most

notably in the domain of painting (i.e., ‘oil paint; oil painting’) and religion (i.e., ‘holy oil’). While enduring coetaneous semantic change, *olio* and *aceite* may never have been identical in meaning (i.e., true synonyms).

As were many other important Arabisms introduced into speech, *aceite* was abundant in the written register, including the scientific output of Alfonso X. The significance of its use in both popular and learned registers is suggestive of a broader acceptance of the term in Castilian speech communities. In part, its acceptance may have been due to its Castilian-like form, which does not contain obvious Arabic phonological, morphological, or syntactic features.

In general, agricultural Arabisms demonstrate a range of formal variation, with some having few documented formal variants and others having many. Since examples such as *alfóncigo*, *arrayán*, and even *aceite* exhibit numerous formal variants, we have found that the factor of formal variation is generally unfavorable to the retention of agricultural Arabisms.

In our examination of 66 agricultural Arabisms found in Maíllo Salgado (1998), we found that most (n = 37) have documented derived terms, whereas others (n = 29) do not. In this way, *aceite* is like most agricultural Arabisms, although its derivative output (n = 10) is greater than other such terms. Again, this is likely due to the greater economic importance of olive oil compared to other agricultural products in the medieval and early modern eras (cf. Martínez Enamorado 2007). That an Arabism (e.g., *aceite*) was introduced despite the preexistence of a partial synonym (i.e., *olio*) is not unknown to Castilian, as almost half of the agricultural Arabisms that we examined overlap in meaning with Latinate or Romance counterparts. In many

cases, however, the introduction may be explained by slight nuances in meaning in the Arabic-Romance lexical items, with a typical outcome being the reservation of the Arabism (e.g., *acebuche* 'wild olive (tree)') for a wild version of the same plant (e.g., *olivo* 'olive (tree)') (Corriente 2008: *acebuche*). In other words, maintenance is supported by functional synonymy.

Aceite differs from other agricultural Arabisms that compete with Romance synonyms such as *alhucema* (vs *espliego ~ lavanda*), *almoraduj* (vs *mejorana*), and *arrayán* (vs *mirto ~ murta*) in that *aceite* is a standard term throughout the Spanish-speaking world, whereas these last three Arabisms are generally restricted to the southernmost varieties of Peninsular Castilian (Maíllo Salgado 1998). With regard to register differences, *aceite* is similar to other Arabisms that are preferred for popular or quotidian language rather than formal or specialized language. As with formal variation, agricultural Arabisms demonstrate a range of acceptance in Castilian, with some terms explicitly condemned (e.g., *alholí, almocafe*; cf. Giménez-Eguíbar 2016) and others (e.g., *aceite*) instead considered an integral part of the language. Acknowledging the range of results demonstrated in formal variation and in attitudes, we may claim that *aceite* is, in many respects, a prototypical agricultural Arabism. However, as we stated above, its usage is not determined by dialectal differences as some such terms are.

6.5 *atalaya* and military Arabisms

In chapter four, we discussed the lexical pair of *atalaya* and *centinela* and other military Arabisms. Addressing linguistic and sociolinguistic factors, we found that semantic differences and the circumstances of language contact, particularly neutral attitudes and intensity of contact, most favored the retention of the term.

We attribute the introduction of *atalaya* to the low frequency of Latin reflexes SPECULA and VIGIL in Castilian Romance (cf. Weinreich 1968). Since these equivalents were poorly established in Castilian Romance, speakers accepted a term (i.e., *atalaya*) while in contact with Hispano-Arabic.

In Castilian Romance, *atalaya* is used to mean both 'watchtower' and the metonymic '(diurnal) sentinel.' The 'watchtower' meaning of the term is relatively stable, while 'sentinel' undergoes semantic generalization. With regard to the 'sentinel' meaning, the term eventually would apply to a sentinel regardless of the time of day he was at his post. Through metaphor, writers would also later use *atalaya* as 'guardian.' As a part of the lexical transition in Castilian, writers borrowed an Italianism, *centinela,* as a more precise (i.e., less ambiguous) term for '(nocturnal) sentinel,' given that *atalaya*, at the time, meant both 'sentinel' and 'watchtower' (cf. Eberenz 2004, López Vallejo 2008). As with *atalaya* 'sentinel,' *centinela* underwent semantic generalization, so that time of day of the post became irrelevant. Also, like its counterpart, *centinela* would be used as 'guardian' through metaphor. With both terms having the general meaning of 'sentinel,' *centinela* displaced the more ambiguous *atalaya* as 'sentinel.'

The circumstances of language contact between Castilian Romance and Arabic are also favorable to *atalaya*. As we have stated, the intensity of contact between Arabic and Castilian Romance during military contact is favorable to the retention of military Arabisms (García González 2007), which were transmitted through speech. Once incorporated into Castilian Romance, awareness of military terms appears to be minimal, as most such Arabisms appear to be undetected as such in the documentation. *Atalaya* in particular could easily be a patrimonial

term based on its formal features alone, and the documentation offers little evidence of its Semitic origin. Where *atalaya* is used metalinguistically, commentary is at least neutral (and, in one case, seemingly positive), which undoubtedly favors its maintenance.

As a military Arabism, *atalaya* has very limited formal variation, especially for one documented as early as the eleventh century, although those that are well established in the twelfth and thirteenth centuries (like *atalaya*) are often less variable. Like several other military Arabisms that are well established in the thirteenth century (e.g., *algara*, *rebato*), *atalaya* is a highly productive lexical base. *Atalaya* is also a functional synonym of a Romance term, something that is true of many, if not most, military Arabisms.

Geographic distribution of *atalaya*, and military Arabisms in a broader sense, is generalized. Also, such terms are generally preferred for popular language, whereas Romance or Latinate equivalents are preferred for formal language. Like other, similar terms, *atalaya* is transmitted through speech as a result of military conflict between Arabs and Christians during the Reconquest (García González 2007). The intensity (i.e., duration) of contact may explain why these terms are generally viewed neutrally. With these facts in mind, *atalaya* may certainly be considered a prototypical military Arabism.

6.6 *jaqueca* and medical Arabisms

In chapter five, we examined the lexical group of *jaqueca, hemicránea,* and *migraña*, as well as other medical Arabisms. Of the linguistic factors examined, formal variation and register are most favorable to the Arabism, whereas the circumstances of language contact (i.e., transmission through speech, neutral attitudes) are the sociolinguistic factor that most favors it.

With regard to formal variation, *jaqueca* presents far greater stability (i.e., in orthography and phonology) than its Latinate counterpart, *hemicránea*. This fact is a reflection of the greater integration of *jaqueca* in the Castilian speech community compared to *hemicránea*, which was likely reserved for specific language domains (i.e., science, medicine) and users (i.e., the learned elite) (Herrera and Vázquez de Benito 1985).

In the documentation, *jaqueca* and *hemicránea* are distinguished by different register usage. *Jaqueca* appears in various text types and is, therefore, a more versatile term, whereas *hemicránea* is restricted to scientific prose. Of the two terms, *jaqueca* is the only one to appear in popular language, where *hemicránea* fails to appear entirely. When there exist two equivalents for medical terms, one tends to be Latin, Greek, or hybrid Latin-Greek, while the other is popular in origin, whether it is a patrimonial term or a borrowing (e.g., *cefalea~dolor de cabeza*); this is certainly the case with *jaqueca* and *hemicránea*.

Another argument that we propose is the perceived utility of a (quotidian) term (i.e., *jaqueca*) for 'migraine' since we did not encounter evidence of a patrimonial term for this specific type of headache. Even though a Latinate term (i.e., *hemicránea*) appears in the documentation, it might only have been accessible to those highly educated in the medical sciences (Herrera and Vázquez de Benito 1985) given its overall infrequency in the documentation, its nearly exclusive usage in medical texts, and the tendency for it to appear in apposition to the more familiar *jaqueca*. In contrast, *jaqueca* is frequent, found in texts of many types, and is usually defined in paraphrase (e.g., *dolor de la media cabeça, dolor de la meytat*

dela cabeça, etc.), rather than being placed in apposition to its Latinate synonym. In other words, with limited access to *hemicránea*, on one hand, and the precision offered by *jaqueca*, on the other (compared to broader OSp. *dolor de cabeça* 'headache'), Hispano-Romance speakers could have incorporated the term, having identified its utility for a specific concept that was familiar, even if a term for it was not. Alternatively, based on its use in Alfonso X, we may also rely on the argument for intensity (i.e., duration) on contact between Arabic and Romance speakers. In either case, *jaqueca*—like the other two Arabisms from our case studies—may also be considered a cultural borrowing.

Since many medical Arabisms were lost from Castilian due to generally negative attitudes toward Arabisms, one might have predicted that *jaqueca* too would have been lost, especially when considering that the Arabism has had to compete (eventualy) with two synonyms. However, in the documentation, we found only neutral commentary toward this particular Arabism, a factor that certainly favors retention.

As a medical Arabism, formal variation of *jaqueca* is moderate, while many such terms (especially those transmitted through writing) are highly variable. Although most medical Arabisms serve as the base for no more than one derived term, *jaqueca* is rather productive, having given rise to at least four different terms. Another feature of *jaqueca* that is unusual is that it coexists with true synonyms; some other retained medical Arabisms are only functional synonyms of Romance or Latinism counterparts (e.g., *alfombrilla*, *jarabe*, *zaratán*).

On the other hand, *jaqueca* is like most medical Arabisms in its generalized geographical distribution. Still, the term differs from others in that it is retained alongside a learned synonym; usually medical Arabisms with Latinate or Romance equivalents that are true synonyms (e.g., *anzarote, cifaque*) are eliminated. One additional feature that distinguishes *jaqueca* from other medical Arabisms is that it was transmitted through speech, where—based on our findings—Arabisms were less subjected to negative attitudes. Since we have found *jaqueca* to be similar to other medical Arabisms with regard to the factor of dialect only (i.e., generalized distribution), we cannot consider the term to be particularly representative of a typical medical Arabism, especially those transmitted through writing.

6.7 The treatment and retention of Arabisms in general[157]

Beyond the treatment of individual Arabisms or even that of the different semantic domains populated by Arabisms, we must also consider what these case studies reveal about the maintenance of Arabisms in general. The case studies suggest that formal variation in and of itself may not have been the decisive factor in the elimination of Arabisms on a large scale. For example, formal variation of *aceite* is considerably greater than that of *olio*. On the other hand, this factor does appear to reveal the degree of integration of terms, which is much greater for certain Arabisms (e.g., *atalaya* and *jaqueca*) over their Romance or Latinism counterparts (e.g., *centinela*, *hemicránea*). In other words, limited formal variation may have a supporting role in the retention of Arabisms.[158]

[157] Our argument for this section is supported by the concept of cultural borrowing, which we discussed in section 2.3.1.
[158] We proposed an argument parallel to that of Cano Aguilar (1993): if excessive formal variation may increase susceptibility to lexical loss, then limited formal variation may help with lexical maintenance.

In a related matter, the degree of integration of the Arabisms that we studied is reflected in the terminology derived from them. After all, derivation is only possible if words are considered a part of the linguistic system in question. We argued that once a semantic (sub)category was well established, it was more likely for the original base to be retained, a reverse but parallel argument to the observation of Dworkin (1989) that once a base word is lost, its derived vocabulary is more susceptible to elimination. In fact, of all the Arabisms that we examined, the three terms that we examined closely are some of the most productive and most successful (i.e., widely and frequently used) in their respective domains, with much of their derived vocabulary considerably more frequent in usage than that of their Latin or Romance counterparts.

Semantic differences between Arabisms and Romance or Latinate counterparts are very favorable to retention in the two cases (i.e., *aceite* 'oil' vs *óleo* 'holy oil; oil paint; oil painting' and *atalaya* 'watchtower; sentinel' vs *centinela* 'sentinel') where it applies. After all, without a true synonymous counterpart to replace the Arabism, it is reasonable to assume that there would be less motivation for its elimination. The maintenance of many Arabisms in agriculture, the military, and medicine that coexist with Latinate or Romance functional (i.e., partial) equivalents could also be explained by these differences. That is, semantic differences—whether due to hyponymy or hypernymy (e.g., *mazmorra* 'dungeon, subterranean prison' ~ *prisión* 'prison'), number of meanings (e.g., *alfóncigo* 'pistachio (tree); pistachio (fruit)' ~ *pistacho* 'pistachio (fruit)'), or connotative differences (e.g., *badea* 'watermelon or melon of low quality' ~ *melón* 'melon')— may explain the retention of Arabisms in a given domain (cf. Grondelaers and Geeraerts 2003).

Although dialectal differences are a factor that has been cited as relevant in the retention of Arabisms (cf. Maíllo Salgado 1998), it is one that did not appear to be relevant for any of our case studies. Although the factor is generally associated with agricultural Arabisms, we can confirm that it is not relevant to the retention of *aceite* itself, but it is for others, as in terminology in the domain of oleiculture in Andalusia. The inclusion of the factor of dialect was due to our curiosity as to whether military and medical Arabisms behaved in the same manner. Although military terms are mostly generalized geographically, there are a few medical Arabisms that are also regional (cf. Garulo 1983; e.g., *alcahuete* 'facial blemish,' *alcatifa* 'head,' *alferecía* 'palsy; paralysis'); however, *jaqueca* is not one of these examples.[159]

A factor that appears to be favorable to many Arabisms, especially those in the domains of agriculture and medicine, is differentiation of usage across register (Escoriza Morera 2002). In general, we found a distinction between the Arabism, on one hand, and its Romance or Latinate equivalent, on the other. In all three case study examples, a Romance or Latinate equivalent is introduced as part of a broader lexical transition in Castilian occurring in the Late Middle Ages and early modern period. Eberenz (2004) argues that motivations for the introduction of new Romance and Latinate borrowings during these eras included greater precision, esthetic purposes, and use for certain written genres, strategies that we agree explain the introduction of *óleo* and *centinela* and the reintroduction or revival of *hemicránea*.[160] All three of these strategies coincide with an intensification of the process of standardization of Castilian. The

[159] In Andalusia, the definitions of these terms differ from those of the standard variety. See Garulo for additional detail and examples (1983: 115-118).
[160] Based on their usage in the documentation, we believe that OSp. *olio* and MSp. *óleo* are distinct terms. Dworkin likewise treats these as separate terms (2012: 95).

somewhat paradoxical result of said lexical transition is that the Romance and Latinate terms, which could have ousted their Arabism counterparts (e.g., *aceite, atalaya, jaqueca*) were not always successful in doing so, perhaps because Arabisms could be used in contexts where more sophisticated vocabulary was unnecessary or less appropriate, such as in casual speech itself or even in writing intended to reflect speech. In other words, in many cases the Latinisms and Romance terms stood in contrast with their Arabism counterparts because the former were essentially distinguished from the latter by register (and, in the case of *aceite* and *atalaya*, semantics).

Based on the considerable success of all three examples that were transmitted through speech and were retained, transmission through speech also appears to favor retention.[161] Since the majority of both agricultural and military terms were transmitted through speech (e.g., *aceite*, *bellota*, *zanahoria*, etc.; *atalaya*, *adalid*, *alferez*, etc.), it is difficult to determine which transmission type is more favorable to the maintenance of Arabisms without considering a domain in which there are numerous examples of both types. Although most medical Arabisms appear to have been transmitted through writing (e.g., *anzarote, cifaque, flisei, jectigacion, mirach, silac*), *jaqueca* and other retained medical Arabisms (e.g., *alcatenes*, *alfombrilla*, *almorí*, *jarabe* / *jarope*, *zaragatona*) were transmitted through speech (cf. Corriente 2008). Based on the elimination of the above indirect medical Arabisms and the retention of the direct medical Arabisms, we are curious to further investigate the matter, namely, to see if direct Arabisms from other domains (e.g., non-medical sciences, musical terms, etc.) are likewise favored over indirect Arabisms. Although we cannot provide evidence for all of the domains in which Arabisms

[161] We defined Arabisms transmitted through speech *direct Arabisms* and those transmitted through writing *indirect Arabisms*.

appear, we do find it reasonable to argue that in the case of medical Arabisms the direct Arabisms were favored in the standard variety over their indirect counterparts due to their tendency to approximate native Castilian forms. Consider the formal variants of three indirect medical Arabisms, which are conspicuously foreign to Castilian: *almirach~mirach*, *flisei~alhasch~ylischi*, *silach~alselach*. In addition, indirect medical Arabisms, although transmitted through writing, tended to have greater formal instability than their direct counterparts, a fact that disfavored retention; after all, one of the purposes of standardization, namely codification, is to minimize variation (Haugen 1972).

Of course, in addition to retention, there is also the consideration of the introduction of Arabisms, which in many cases is encouraged by intensity (i.e., duration) of contact and lexical vacancies in the target language (i.e., Castilian Romance or Castilian). As we stated earlier, loans that result from vacancies in the lexicon are often identified as cultural, or necessary, borrowings since there is no designation for the words in question (cf. Goddard 1969, Dworkin 2012). In general, we consider Arabisms from all three domains to be cultural borrowings, acknowledging that the motivations specific to individual Arabisms are not necessarily the same for all Arabisms in a given domain. In other words, regardless of the specific motivation (e.g., low frequency of native terms, lexical vacancies) for individual borrowings, most Arabisms can be considered cultural borrowings. We agree with García González (1993-1994) that the traditional notion of the prestige of Arabic or Arabisms—even military Arabisms, which are generally viewed neutrally—is untenable. The documentation shows that military borrowings, even military Gallicisms and Italianisms, were condemned by Castilian writers upon their introduction, so to claim that prestige is the primary motivation for the introduction of military Arabisms into

Castilian is misguided. For example, even though the incorporation of military Gallicisms was intense in the eighteenth century (through the translation of military manuals; e.g., *Diccionario militar* and others), the phenomenon was met with a purist reaction associated with the process of standardization (Sánchez Orense 2013). The same may be said of Italianisms: despite their obvious influence in the domain of the military (sixteenth and seventeenth centuries), they are condemned in the documentation by Castilian writers reacting to the notable influx of foreign vocabulary. That said, the negative reaction of some authors toward military Gallicisms and Italianisms does not discount the fact that certain groups viewed these neutrally or even favorably. On the other hand, Arabisms appear to have been viewed neutrally only under the best circumstances, which is why we do not consider Arabisms from any of the three domains to be luxury borrowings. As García González (1993-1994, 2012) points out, the problem with ascribing the use of Arabisms to prestige is doubtful given the many examples of semantic change and pejoration that Arabisms underwent during their passage from Arabic to Hispano-Romance and Castilian.

6.8 Limitations and future research

One limitation of our study is the lack of documentation from non-Hispano-Romance or non-Castilian sources. In particular, a study similar to our own could include Mozarabic (i.e., Andalusi Romance) documentation and/or a greater amount of Arabic language documentation than examined in the current study. In doing so, the researcher could corroborate or identify with greater precision the dating and usage of the Arabisms in question prior to their documentation in Hispano-Romance or Castilian.

Another limitation of this dissertation is our exclusion of quantitative analysis in the project, even though there is potential for the use of applications like GoldVarb or R to address issues in historical linguistics, just as there is in variationist sociolinguistics. We believe that quantitative data could further enhance the value of this project and others like it in the still relatively new discipline of historical sociolinguistics.

Lastly, we acknowledge that our project did not (and could not reasonably) address all examples of Arabisms in the three domains discussed. We instead focused our discussion on case studies to explain the treatment and behavior of Arabisms in general by examining different types of borrowings. However, we must note that there is still much to be learned from other Arabisms and lexical rivalries in agriculture, the military, and medicine, since each word has its own unique history.

Although we are now aware of the factors that most influence the retention of Arabisms from the domains of agriculture, the military, and health, we would be interested in how our study may shed light on the treatment of Arabisms from other domains. As we stated in our introduction, there are numerous other domains where Arabisms were borrowed and maintained, some of which include the following: urbanism and housing (*arrabal* 'suburb', *barrio* 'district, quarter'), domestic life (*almohada* 'pillow,' *zapato* 'shoe'), animals (*alazán* 'sorrell (horse),' *atún* 'tuna'), mineral products (*azogue* 'mercury,' *azufre* 'sulfur'), musical terms (*guitarra* 'guitar,' *tambor* 'drum'), and colors (*azul* 'blue,' *escarlata* 'scarlet'). Giménez-Eguíbar (2011) studied loss of Arabisms in the trades or professions (*alfayate*, *alfajeme*, etc.) in competition; another researcher could address Arabisms in the trades (if any) that maintained their dominance

over Latinate or Romance synonyms. Still, since most of the domains in which Arabisms appear are popular (i.e., not learned) terms, it is likely that many of the same factors that we analyzed will favor the maintenance of Arabisms within these domains as well.

Another study may examine the treatment of Arabic loanwords introduced into Castilian through other Romance languages. We chose to limit our study of Arabisms to words that passed into Castilian Romance or Castilian from Arabic or Hispano-Arabic (rather than through an intermediate language) (cf. Corriente 2012). For example, some maritime terms, such as *almadía* 'raft' and *monzón* 'monsoon' appear to have entered into Castilian through Portuguese. Another indirect entry is that of *sirope* 'syrup,' first borrowed into French before its borrowing into Castilian. We expect that these words of ultimately Arabic origin are treated as other semantically related Romance terms due to their entry from another Romance language, rather than Arabic or Hispano-Arabic since the circumstances of contact that resulted in the borrowing of these terms occurred between Castilian and Portuguese or Castilian and French, which are necessarily different from the contact between Castilian and varieties of Arabic.

One final area of interest for future research is the treatment of Arabisms that are introduced into Castilian from Modern Arabic (e.g., *majzén*, *jalifa*; cf. Morgenthaler García 2014). Although these are far fewer than Arabisms incorporated during the medieval and early modern ages, these new Arabisms also include military terminology. With this in mind, we are curious to discover if the same factors that favor military Arabisms from the Middle Ages and early modern period also favor those from the modern period. Specifically, it may be of interest to compare attitudes toward Arabisms from the past with those of the present. For example, we

are interested in discovering if there are sources of resistance to new Arabisms and, if so, how they compare with those concerning older Arabisms.

References

Bibliography

Agüera Pedrosa, Santos, Antonio Luis Molina Berbel and Andrés Carrillo Miras. 2017. Las torres-atalayas: el sistema defensivo naserita del valle del Almanzora. In Díaz López, Julián Pablo (ed.). 2017. *La historia de Almería y sus historiadores: centenario del padre Tapia,* 301-316. Almería: Instituto de Estudios Almerienses.

Alcalá, Pedro de. 1505. *Arte para ligeramente saber la lengua arauiga*. Granada: Juan Varela de Salamanca.

Aléong, Stanley. 1981. Lexical Variation in Quebec French Ice Hockey Terminology: A Historical Perspective. In Sankoff, David and Henrietta Cedergren (eds.). 1981. *Variation Omnibus*, 463-470. Edmonton: Linguistic Research.

Almeida Suárez, Manuel Ventura. 1994. Creencias y actitudes lingüísticas en el español canario. *Anuario de Lingüística Hispánica* 10: 9-24.

Alvar, Manuel. 1973. *Estructuralismo, geografía lingüística y dialectología actual*. 2nd ed. Madrid: Gredos.

Álvarez de Miranda, Pedro. 2009. Neología y pérdida léxica. In Miguel, Elena de (ed.). 2009. *Panorama de la lexicología*, 189-215. Barcelona: Ariel.

Blas Arroyo, José Luis. 2008. The variable expression of future tense in Peninsular Spanish: the present (and future) of inflectional forms in the Spanish spoken in a bilingual region. *Language Variation and Change* 20 (1): 85-126.

Blas Arroyo, José Luis (ed.). 2009a. *Corpus sociolingüístico de Castellón de la Plana y su área metropolitana*. Castelló de la Plana: Universitat Jaume I. Cited in Blas Arroyo, José Luis. 2009b. La variación léxica. In Miguel, Elena de (ed.). 2009. *Panorama de la lexicología*, 189-215. Barcelona: Ariel.

Blas Arroyo, José Luis. 2009b. La variación léxica. In Miguel, Elena de (ed.). 2009. *Panorama de la lexicología*, 189-215. Barcelona: Ariel.

Blas Arroyo, José Luis. 2012. *Sociolingüística del español: Desarrollos y perspectivas en el estudio de la lengua española en contexto social*. 3rd ed. Madrid: Cátedra.

Boberg, Charles. 2004. Real and Apparent Time in Language Change: Late Adoption of Changes in Montreal English. *American Speech* 79 (3): 250-269.

Boberg, Charles. 2005. The North American Regional Vocabulary Survey: New Variables and Methods in the Study of North American English. *American Speech* 80 (1): 22-60.

Borrego Nieto, Julio. 1981. *Sociolingüística rural: Investigación en Villadepera de Sayago*. Salamanca: Universidad de Salamanca.

Borrego Nieto, Julio. 1994. Dificultades para el estudio sociolingüístico del léxico. In Alonso González, Alegría, Beatriz Garza Cuarón and José Pascual (eds.). 1994. *II Encuentro de Lingüistas y Filólogos de España y México*, 119-131. Salamanca: Junta de Castilla y León, Consejería de Cultura y Turismo.

Bossong, Georg. 1982. Las traducciones alfonsíes y el desarrollo de la prosa científica. In Hempel, Wido and Dietrich Briesemeister (eds.). 1982. *Actas del coloquio hispano-alemán Ramón Menéndez Pidal*, 1-11. Tübingen: Max Niemeyer.

Bréal, Michel. 1964. *Semantics: Studies in the Science of Meaning*. Translation by Nina Cust. New York: Dover.

Burns, Robert. 1967. *The Crusader Kingdom of Valencia*. Cambridge: Harvard University.
Butt, John and Carmen Benjamin. 2004. *A New Reference Grammar of Modern Spanish*. 4th ed. New York: McGraw-Hill.

Cabañero Subiza, Bernabé. 1996. *Los castillos catalanes del siglo X: circunstancias históricas y cuestiones arquitectónicas*. Zaragoza: Institución Fernando el Católico.

Cano Aguilar, Rafael. 1993. La muerte de las palabras. In Alcaide Lara, Esperanza, María del Mar Ramos Márquez and Francisco Salguero Lamillar (eds.). 1993. *Estudios lingüísticos en torno a la palabra*, 41-57. Seville: Universidad de Sevilla.

Carrasco Cantos, Inés. 1992. *Los cargos de la hueste real en tiempos de Alfonso X: estudio onomasiológico*. Granada: Universidad de Granada.

Casado Alonso, Hilario and Teofilo Ruiz. 2019. Economic Life in Late Medieval and Early Modern Spain, 1085–1815. *Oxford Research Encyclopedia of Economics and Finance*. Available online: https://oxfordre.com/economics/view/10.1093/acrefore/9780190625979.001.0001/acrefore-9780190625979-e-481 (Consulted 12 May 2020).

Castro, Américo. 1922. *Ojo* < OLEUM, *nidio* < NITIDUM, *lezne* < LICINUM. *Revista de Filología Española* 9: 65-67.

Chandler, William. 1950. *Evergreen Orchards*. Philadelphia: Lea and Febiger.

Chejne, Anwar. 1974. *Muslim Spain: Its History and Culture*. Minneapolis: University of Minnesota.

Colón, Germán. 1967. Occitanismos. In Alvar, Manuel, Antonio Badía Margarit, Rafael de Balbín, Luís Filipe Lindley Cintra. 1967. *Enciclopedia lingüística hispánica*. Vol. 2, *Elementos constitutivos*. Madrid: Consejo Superior de Investigaciones Científicas.

Colón, Germán. 1998. Para una presentación del léxico español. *Acta Romanica Basiliensia* 8: 215-230.

Corominas, Joan and José Pascual. 1980-1991. *Diccionario crítico etimológico castellano e hispánico.* Madrid: Gredos.

Corriente, Federico. 2005. El elemento árabe en la historia lingüística peninsular: Actuación directa e indirecta. Los arabismos en los romances peninsulares (en especial en castellano). In Cano Aguilar, Rafael (ed.). 2005. *Historia de la lengua española*, 185-204. Barcelona: Ariel.

Corriente, Federico. 2008. *Dictionary of Arabic and Allied Loanwords: Spanish, Portuguese, Catalan, Galician and Kindred Dialects*. Boston: Brill.

Corriente, Federico. 2012. Arabismos lusismos del español de Canarias. In Corbella, Dolores, Josefa Dorta and Alejandro Fajardo Aguirre (eds.). 2012. *Lexicografía hispánica del siglo XXI: Nuevos proyectos y perspectivas: Homenaje al Profesor Cristóbal Corrales Zumbado*, 211-218. Madrid: Arco Libros.

Craddock, Jerry. 1990. The Legislative Works of Alfonso el Sabio. In Burns, Robert (ed.). 1990. *Emperor of Culture: Alfonso X the Learned of Castile and His Thirteenth-Century Renaissance*, 182-197. Philadelphia: University of Pennsylvania.

Cressier, Patrice. 2004. Vigilar, proteger, habitar: a propósito de las torres en el ámbito rural andalusí. In Moret, Pierre and Teresa Chapa (eds.). 2004. *Torres, atalayas y casas fortificadas: explotación y control del territorio en Hispania (s. III a. de C. - s. I d. de C.)*, 209-213. Jaén: Universidad de Jaén.

Dufourcq, Charles Emmanuel and Jean Gautier Dalché. 1983. *Historia económica y social de la España cristiana en la Edad Media*. Barcelona: El Albis.

Dworkin, Steven. 1989. Factores lingüísticos operantes en la pérdida léxica. In Kremer, Dieter (ed.). 1989. *Actes du XVIII Congrès International de Linguistique et Philologie Romanes* 4, 379-384. Tübingen: Max Niemeyer Verlag.

Dworkin, Steven. 2004. La transición léxica en el español bajomedieval. In Cano Aguilar, Rafael (ed.). 2004. *Historia de la lengua española*. Barcelona: Ariel.

Dworkin, Steven. 2006. La historia de la lengua y el cambio léxico. *Iberoromania* 62 (2): 59-70.

Dworkin, Steven. 2012. *A History of the Spanish Lexicon*. New York: Oxford University.

Eberenz, Rolf. 2004. En torno al léxico fundamental del siglo XV: sobre algunos cambios verbales. In Lüdtke, Jens and Christian Schmitt (eds.). 2004. *Historia del léxico español: enfoques y aplicaciones*, 111-136. Madrid: Iberoamericana.

Eberenz, Rolf. 2006. Cultura lingüística y cultivo del castellano en el otoño de la Edad Media. In Bustos Tovar, José Jesús de and José Luis Girón Alconchel (eds.). 2006. *Actas del VI Congreso Internacional de Historia de la Lengua Española* 1, 85-102. Madrid: Arco Libros.

Escoriza Morera, Luis. 2002. *La variación lingüística: Propuesta de delimitación de variantes en el nivel léxico*. Doctoral thesis. Cádiz: Universidad de Cádiz.

Etxebarría, Maitena. 1985. *Sociolingüística urbana: El habla de Bilbao*. Salamanca: Universidad de Salamanca.

Fernández-Sevilla Jiménez, Julio. 1975. *Formas y estructuras en el léxico agrícola andaluz*. Madrid: Consejo Superior de Investigaciones Científicas.

Fernández Smith, Gérard and Luis Escoriza Morera. 2004. Variación léxica y texto: Análisis cualitativo del grado de formalidad en contexto con AQUAD. *Pragmalingüística* 12: 73-90.

Fries, Charles and Kenneth Pike. 1949. Coexistent phonemic systems. *Language* 25: 29-50.

García de Cortázar, José Ángel. 1982. *La formación de la sociedad hispano-cristiana del Cantábrico al Ebro en los siglos VIII a XI*. Santander: Librería Estudio.

García González, Javier. 1993-1994. El contacto de dos lenguas: Los arabismos en el español medieval y en la obra alfonsí. *Cahiers de Linguistique Hispanique Médievale* 18-19: 335-365.

García González, Javier. 1996-1997. Los arabismos militares y los referentes a lo árabe en los textos alfonsíes. *Cahiers de Linguistique Hispanique Médiévale* 21: 127-144.

García González, Javier. 2007. Una perspectiva sociolingüística de los arabismos en el español de la alta Edad Media (711-1300). In Puigvert Ocal, Alicia and Inmaculada Delgado Cobos (eds.). 2007. *Ex admiratione et amicitia: Homenaje a Ramón Santiago* 1, 523-548. Madrid: Ediciones del Orto.

García González, Javier and María Luisa Coronado González. 1998. ¿Qué lengua enseñamos a nuestros alumnos extranjeros? Algunas peculiaridades del lenguaje femenino. In Fente Gómez, Rafael, Antonio Martínez González and José Andrés de Molina Redondo (eds.). 1998. *El español como lengua extranjera: Aspectos generales*, 119-128. Málaga: Asociación para la Enseñanza del Español como Lengua Extranjera.

García Mouton, Pilar. 1996. Dialectología y geografía lingüística. In Alvar, Manuel (ed.). 1996. *Manual de dialectología hispánica: El español de España*, 63-77. Barcelona: Ariel.

Garulo, Teresa. 1983. *Los arabismos en el léxico andaluz según los datos del Atlas Lingüístico y Etnográfico de Andalucía*. Madrid: Instituto Hispano-Árabe de Cultura.

Geeraerts, Dirk, Stefan Grondelaers and Peter Bakema. 1994. *The Structure of Lexical Variation: Meaning, Naming and Context*. Berlin: Mouton de Gruyter.

Gil Crespo, Ignacio Javier. 2013. *Fundamentos constructivos de las fortificaciones fronterizas entre las coronas de Castilla y Aragón de los siglos XII al XV en la actual provincia de Soria*. Doctoral thesis. Madrid: Universidad Politécnica de Madrid.

Gilman, Charles. 1979. Lexical Loss in Cameron Pidgin. *Anthropos* 74: 174-180.

Giménez-Eguíbar, Patricia. 2011. *Arabismos en el campo semántico de los oficios: De la competición a la perdida lexica*. Doctoral thesis. Madison: University of Wisconsin.

Giménez-Eguíbar, Patricia. 2016. Attitudes Towards Lexical Arabisms in Sixteenth-Century Spanish Texts. In Sessarego, Sandro and Fernando Tejedo-Herrero (eds.). 2016. *Spanish Language and Sociolinguistic Analysis*, 363-380. Amsterdam; Philadelphia: John Benjamins.

Goddard, K.A. 1969. Loan-words and lexical borrowing in Romance. *Revue de linguistique romane* 33: 337-348.

Gómez-Moreno Martínez, Manuel. 1919. *Iglesias mozárabes*. Madrid: Centro de Estudios Históricos.

Grondelaers, Stefan and Dirk Geeraerts. 2003. Towards a Pragmatic Model of Cognitive Onomasiology. In Cuyckens, Hubert, René Dirven and John R. Taylor (eds.). 2003. *Cognitive Approaches to Lexical Semantics*, 67-92. Berlin: Mouton de Gruyter.

Grondelaers, Stefan, Dirk Geeraerts and José Tummers. 2001. Lexical standardisation in internet conversations. Comparing Belgium and The Netherlands. In Fontana, Josep M., Louise McNally, María Teresa Turell and Enric Vallduví (eds.). 2001. *Proceedings of the First International Conference on Language Variation in Europe*, 90-100. Barcelona: Universitat Pompeu Fabra.

Gutiérrez Rodilla, Bertha. 2005. *El lenguaje de las ciencias*. Madrid: Gredos.

Harris-Northall, Ray. 1999. Re-Latinization of Castilian Lexis in the Early Sixteenth Century. *Bulletin of Hispanic Studies* 76: 1-12.

Haugen, Einar. 1972. Dialect, Language, Nation. In Pride, John B. and Janet Holmes (eds.). 1972. *Sociolinguistics: Selected Readings*, 97-111. Harmondsworth: Penguin.

Herrera, María Teresa. 1996. *Diccionario español de textos médicos antiguos*. Madrid: Arco/Libros.

Herrera, María Teresa, María de las Nieves Sánchez González de Herrero and María Purificación Zabía. 1997. *Términos definidos en el corpus base del* DETEMA. Madison: Hispanic Seminary of Medieval Studies.

Herrera, María Teresa and María Concepción Vázquez de Benito. 1981. Arabismos en el castellano de la medicina y farmacopea medievales: apuntes para un nuevo diccionario (I). *Cahiers de linguistique hispanique médiévale* 6: 123-169.

Herrera, María Teresa and María Concepción Vázquez de Benito. 1982. Arabismos en el castellano de la medicina y farmacopea medievales: apuntes para un nuevo diccionario (II). *Cahiers de linguistique hispanique médiévale* 7 (1): 173-216.

Herrera, María Teresa and María Concepción Vázquez de Benito. 1983. Arabismos en el castellano de la medicina y farmacopea medievales: apuntes para un nuevo diccionario (III). *Cahiers de linguistique hispanique médiévale* 8: 165-196.

Herrera, María Teresa and María Concepción Vázquez de Benito. 1985. Arabismos en el castellano de la medicina y farmacopea medievales: apuntes para un nuevo diccionario (IV). *Cahiers de linguistique hispanique médiévale* 10: 71-100.

Hitti, Philip. 1951. *History of the Arabs*. London: Macmillan.

Holmes, Janet. 1992. *An Introduction to Sociolinguistics*. London: Longman.

Janda, Richard and Brian Joseph. 2003. On Language, Change, and Language Change – Or, Of History, Linguistics, and Historical Linguistics. In Joseph, Brian and Richard Janda (eds.). 2003. *The Handbook of Historical Linguistics*, 3-180. Malden, MA: Blackwell.

Kasten, Lloyd. 1990. Alfonso el Sabio and the Thirteenth-Century Spanish Language. In Burns, Robert (ed.). 1990. *Emperor of Culture: Alfonso X the Learned of Castile and His Thirteenth-Century Renaissance*, 33-45. Philadelphia: University of Pennsylvania.

Knaus, Valerie and Terry Nadasdi. 2001. *Être* ou ne pas *être* in Immersion French. *The Canadian Modern Language Review* 58 (2): 287-306.

Koerner, Konrad. 1991. Toward a history of sociolinguistics. *American Speech* 66 (1): 57-70.

Laberge, Suzanne. 1978. *Étude de la variation des pronoms sujets définis et indéfinis dans le français parlé à Montréal*. Doctoral thesis. Montreal: Université de Montréal.

Labov, William. 1966. *The Social Stratification of English in New York City*. Washington: Center for Applied Linguistics.

Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania.

Labov, William. 1978. Where Does the Linguistic Variable Stop? A Response to Beatriz Lavandera. *Working Papers in Sociolinguistics* 44: 1-22.

Labov, William. 1984. Field Methods of the Project on Linguistic Change and Variation. In Baugh, John and Joel Sherzer (eds.). 1984. *Language in Use: Readings in Sociolinguistics*, 28-54. Englewood Cliffs, NJ: Prentice-Hall.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Cambridge, MA: Blackwell.

Lampérez y Romea, Vicente. 1922. *Arquitectura civil española de los siglos I al XVIII*. Madrid: Saturnino Calleja.

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

Lappalainen, Hanna. 2004. Variation and Its Functions. The Analysis of Linguistic Variation and Interaction among Members of a Social Network. In Gunnarsson, Britt-Louise (ed.). 2004. *Language Variation in Europe*, 267-280. Uppsala: Uppsala Universitet.

Lavandera, Beatriz. 1978. Where Does the Sociolinguistic Variable Stop? *Language in Society* 7 (2): 171-182.

Lipski, John. 2010. Spanish and Portuguese in Contact. In Hickey, Raymond (ed.). 2010. *The Handbook of Language Contact*, 550-580. Malden, MA: Blackwell.

Lodares, Juan Ramón. 1992. Lexicología histórica e historia social. In Ariza Viguera, Manuel (ed.). 1992. *Actas del II Congreso Internacional de Historia de la Lengua Española* 1, 1145-1150. Madrid: Pabellón de España.

López Morales, Humberto. 2004. *Sociolingüística*. Madrid: Gredos.

López Morales, Humberto. 2005. Sociolingüística del tabú. *Interlingüística* 16 (1): 7-20.

López Vallejo, María Ángeles. 2006. Italianismos militares en el Renacimiento: la palabra *centinela*, un ejemplo. In Rodríguez Molina, Javier and Daniel Sáez Rivera (eds.). 2006. *Diacronía, lengua española y lingüística: Actas del IV Congreso Nacional de la Asociación de Jóvenes Investigadores de Historiografía e Historia de la Lengua Española*, 355-368. Madrid: Síntesis.

López Vallejo, María Ángeles. 2008. *Historia del léxico militar en el español áureo: la conquista de Granada, el conflicto hispano-italiano y las guerras de Flandes*. Doctoral thesis. Granada: Universidad de Granada.

Maíllo Salgado, Felipe. 1998. *Los arabismos del castellano en la Baja Edad Media*. Salamanca: Universidad de Salamanca.

Maíllo Salgado, Felipe. 2004. *De la desaparición de al-Andalus*. Madrid: Abada.

Malkiel, Yakov. 1976. La fragmentación de FĪDŪCIA en hispanorrománico. *Nueva Revista de Filología Hispánica* 26 (1): 1-55.

Martínez Enamorado, Virgilio. 2007. 'Aceite que viene de tierra de moros': algunos datos sobre la comercialización del aceite de oliva andalusí. In VV.AA. 2007. *Tierras del olivo*, 169-175. Sevilla: Fundación El Legado Andalusí.

McMahon, April. 1995. *Understanding Language Change*. Cambridge: Cambridge University.

Meillet, Antoine. 1912. L'évolution des formes grammaticales. In *Linguistique historique et linguistique générale*. Paris: Champion.

Moreno Fernández, Francisco. 1999. Análisis cuantitativo de campos léxicos. In Blecua, José Manuel (ed.). 1999. *Filología e informática: Nuevas tecnologías en los estudios filológicos*, 357-388. Barcelona: Universidad Autónoma de Barcelona.

Moreno Fernández, Francisco. 2005. *Historia social de las lenguas de España*. Barcelona: Ariel.

Morgenthaler García, Laura. 2014. Nuevos arabismos en el español: Una revisión diacrónica y crítica a partir del siglo XIX. In Zimmermann, Klaus (ed.). 2014. *Prácticas y políticas lingüísticas: Nuevas variedades, normas, actitudes y perspectivas*, 315-337. Madrid: Iberoamericana.

Mougeon, Raymond and Édouard Beniak. 1991. *The Linguistic Consequences of Language Contact and Restriction: The Case of French in Ontario, Canada*. Oxford: Oxford University.

Mougeon, Raymond and Katherine Rehner. 2001. Acquisition of Sociolinguistic Variants by French Immersion Students: The Case of Restrictive Expressions, and More. *Modern Language Journal* 85 (3): 398-415.

Muñoz Núñez, María Dolores. 1999. *La polisemia léxica*. Cádiz: Universidad de Cádiz.

Nadasdi, Terry. 2005. Living in Ontario French. *Canadian Journal of Applied Linguistics* 8 (2): 167-181.

Nadasdi, Terry and Meghan McKinnie. 2003. Living and working in immersion French. *Journal of French Language Studies* 13 (1): 47-61.

Nebrija, Antonio de. 1980. *Gramática de la lengua castellana*. Madrid: Editora Nacional.

Neuvonen, Eero. 1941. Los arabismos del español en el siglo XIII. *Studia Orientalia* 10: 2-332.

Penny, Ralph. 2001. *Variation and Change in Spanish*. Cambridge: Cambridge University.

Penny, Ralph. 2002. *A History of the Spanish Language*. Cambridge: Cambridge University.

Pensado Figueiras, Jesús. 2012. *El códice Zabálburu de medicina medieval: edición crítica y estudio de fuentes*. Doctoral thesis. A Coruña: Universidade da Coruña.

Pérez Vidal, Carmelo and Marina Díaz Peralta. 1996. Actitudes sociolingüísticas ante el léxico en español actual. In Arjona Iglesias, Marina (ed.). 1996. *Actas del X Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina*, 833-837. Mexico City: Universidad Nacional Autónoma de México.

Picornell Buendía, María Raquel and José María Melero Martínez. 2013. Historia del cultivo del olivo y del aceite; su expresión en la Biblia. *ENSAYOS: Revista de la Facultad de Educación de Albacete* 28: 155-181.

Read, Jan. 1974. *The Moors in Spain and Portugal*. London: Faber and Faber.

Ridruejo Alonso, Emilio. 1998. La inserción de -g- en el presente de caigo, oigo, traigo. In García Turza, Claudio, Fabián González Bachiller, José Javier Mangado Martínez (eds.). 1998. *Actas del IV Congreso Internacional de Historia de la Lengua Española* 1, 725-734. Logroño: Universidad de La Rioja.

Rudskoger, Arne. 1952. *Fair, Foul, Nice, Proper: A Contribution to the Study of Polysemy*. Stockholm: Almquist and Wiksell.

Sánchez González de Herrero, María de las Nieves. 1998. Características lingüísticas del "Tratado de Patología" (siglo XV). In García Turza, Claudio, Fabián González Bachiller and José Javier Mangado Martínez (eds.). 1998. *Actas del IV Congreso Internacional de Historia de la Lengua Española* 2, 847-856. Logroño: Universidad de La Rioja.

Sánchez González de Herrero, María de las Nieves and María Concepción Vázquez de Benito. 2010. La traducción de textos médicos medievales: cuestiones en torno al léxico. *Cuadernos del Instituto Historia de la Lengua* 4: 77-103.

Sánchez Orense, Marta. 2013. El *Diccionario militar* (1749) de Raimundo Sanz y la incorporación de galicismos al castellano. *Anuari de filologia. Estudis de lingüística* 3: 159-183.

Sankoff, Gillian. 1972. Above and Beyond Phonology in Variable Rules. In Bailey, Charles-James and Roger Shuy (eds.). 1972. *New Ways of Analyzing Variation in English*, 44-61. Washington, D.C.: Georgetown University.

Sankoff, Gillian and Paul Thibault. 1977. L'alternance entre les auxiliaires *avoir* et *être* en français parlé à Montréal. *Langue française* 34: 81-108.

Sapir, Edward. 1921. *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and Company.

Sedano, Mercedes. 1994. Evaluation of Two Hypotheses about the Alternation between *aquí* and *acá* in a Corpus of Present-Day Spanish. *Language Variation and Change* 6 (2): 223-237.

Tagliamonte, Sali A. 2006. *Analysing Sociolinguistic Variation*. New York: Cambridge University.

Tannahill, Reay. 1989. *Food in History*. New York: Three Rivers.

Terlingen, Johannes. 2016. *Los italianismos en español: desde la formación del idioma hasta principios del siglo XVII*. Seville: Athenaica.

Thomason, Sarah. 2001. Language Contact: An Introduction. Edinburgh: Edinburgh University.

Traugott, Elizabeth and Richard Dasher. 2003. *Regularity in Semantic Change*. Cambridge; New York: Cambridge University.

Trillo San José, Carmen. 2007. El olivo en al-Andalus: tradición latina e islámica. In VV.AA. 2007. *Tierras del olivo*, 105-115. Sevilla: Fundación El Legado Andalusí.

Tuten, Donald. 2003. *Koineization in Medieval Spanish*. Berlin; New York: Mouton de Gruyter.

Tuten, Donald and Fernando Tejedo-Herrero. 2011. The Relationship between Historical Linguistics and Sociolinguistics. In Díaz-Campos, Manuel (ed.). 2011. *The Handbook of Hispanic Sociolinguistics*, 283-302. Malden, MA: Blackwell.

Ullmann, Stephen. 1962. *Semantics: An Introduction to the Science of Meaning*. Oxford: Blackwell.

Ullmann, Stephen. 1964. *Language and Style*. Oxford: Blackwell.

Vaca Lorenzo, Ángel. 2016. Expansión agraria, urbana y comercial en los siglos XI al XII. In Monsalvo Antón, José María (ed.). 2016. *Historia de la España medieval*, 185-220. Salamanca: Universidad de Salamanca.

Vaquerizo Gil, Desiderio. 2011. De especie silvestre (*kótinos*) a olivo sagrado (*élaios*): notas sobre el cultivo del olivar, la producción, la comercialización y el consumo de aceite de oliva en el Mediterráneo antiguo. *Boletín de la Real Academia de Extremadura de las Letras y las Artes* 19: 591-673.

Vicens Vives, Jaime. 1969. *An Economic History of Spain*. Princeton: Princeton University.

Weinreich, Uriel. 1968. *Languages in Contact: Findings and Problems*. The Hague: Mouton de Gruyter.

Weinreich, Uriel, William Labov and Marvin Herzog. 1968. Empirical Foundations for a Theory of Language Change. In Lehmann, Winfred and Yakov Malkiel (eds.). 1968. *Directions for Historical Linguistics*, 95-105. Austin: University of Texas.

Winet, Monika. 2006. *El artículo árabe en las lenguas iberorrománicas: Aspectos fonéticos, morfológicos y semánticos de la transferencia léxica*. Córdoba: Universidad de Córdoba.

Zahran, Rabie. 2006. *Sistemas defensivos y técnicas constructivas en el sector accitano del Reino de Granada*. Doctoral thesis. Granada: Universidad de Granada.

Corpora and dictionaries

Merriam-Webster, Inc. 2018. *Dictionary by Merriam-Webster*. Available online at http://www.merriam-webster.com.

Real Academia Española. CORDE: *Corpus diacrónico del español*. Available online at http://corpus.rae.es/cordenet.html.

Real Academia Española. 2016. CORPES XXI: *Corpus del Español del Siglo XXI*. Available online at http://web.frl.es/CORPES/view/inicioExterno.view.

Real Academia Española. 2008. CREA: *Corpus de referencia del español actual*. Available online at http://corpus.rae.es/creanet.html.

Real Academia Española. 1726-1739. *Diccionario de autoridades*. Available online at http://web.frl.es/DA.html.

Real Academia Española. 2017. *DLE*: *Diccionario de la lengua española*. Available online at http://dle.rae.es/?w=diccionario.

Real Academia Española. *NTLLE: Nuevo tesoro lexicográfico de la lengua española*. Available online at http://ntlle.rae.es/ntlle/SrvltGUILoginNtlle.

Real Academia de la Historia. *Diccionario Biográfico electrónico.* Available online at *http://dbe.rah.es/*.